UNIVERSIDAD AUTÓNOMA DE MADRID

DEPARTMENT OF MOLECULAR BIOLOGY

PhD Thesis

Whole exome sequencing of Urothelial Bladder Cancer: identification of ARID1A and STAG2 as new, important, players

CRISTINA BALBÁS MARTÍNEZ

MADRID, 2014

AUTONOMOUS UNIVERSITY OF MADRID

FACULTY OF SCIENCES

DEPARTMENT OF MOLECULAR BIOLOGY

Whole exome sequencing of Urothelial Bladder Cancer: identification of ARID1A and STAG2 as new, important, players

Doctoral thesis submitted to the Autonomous University of Madrid

for the degree of Doctor of Philosophy by

MSc in Molecular Biomedicine,

Cristina Balbás Martínez

Thesis Director

Francisco X. Real Arribas

EPITHELIAL CARCINOGENESIS GROUP

BBVA FOUNDATION – CNIO CANCER CELL BIOLOGY PROGRAMME

SPANISH NATIONAL CANCER RESEARCH CENTRE

Francisco X. Real Arribas, Head of the Epithelial Carcinogenesis Group at the Spanish National Cancer Research Center (CNIO)

CERTIFIES

That Ms. Cristina Balbás Martínez, Master in Molecular Biomedicine by the Universidad Autónoma de Madrid, has completed her Doctoral Thesis "Whole exome sequencing of Urothelial Bladder Cancer: identification of ARID1A and STAG2 as new, important, players" and meets the necessary requirements to obtain the PhD in Molecular Biosciences. To this purpose, she will defend her Doctoral Thesis at the Universidad Autónoma de Madrid. The work has been carried out under my supervision and hereby I authorize its defence.

I hereby issue this certificate in Madrid on September 1st 2014

Francisco X. Real
PhD Thesis Director

María Jesús Bullido
PhD Thesis Tutor

This thesis, submitted for the degree of Doctor of Philosophy at the Autonomous University of Madrid, has been carried out and completed in the Epithelial Carcinogenesis Group at the Spanish National Cancer Research Centre (CNIO), under the supervision of Francisco X. Real Arribas

This doctoral thesis has been supported by La Caixa through a CNIO International PhD Fellowship. The work was supported by grants from Ministerio de Economía y Competitividad, Madrid, Spain (grants Consolíder ONCOBIO, Consolider INESGEN, SAF-2010-21517 and SAF2011-15934-E), Instituto de Salud Carlos III (grants G03/174, 00/0745, PI051436, PI061614, G03/174, PI080440, PI120425 and Red Temática de Investigación Cooperativa en Cáncer (RTICC)), Asociación Española Contra el Cáncer, EU-FP7-201663 and 201333, and US National Institutes of Health grant RO1 CA089715.

There is more to us than we know. If we can be made to see it, perhaps for the rest of our lives we will be unwilling to settle for less.

Kurt Hahn

No se puede caminar en dos direcciones distintas, pero la gracia de la vida es poder ir a donde tiene que irse por diferentes caminos. Y por la ciencia, como por el arte, se va al mismo sitio: a la verdad. Además, lo que importa es el camino. El camino es el que hace entretenidos los días y gratas las noches. El fin es siempre un sueño. Y quizá el verdadero fin es nunca llegar.

Gregorio Marañón

A mis abuelos: Ángel, Antonio, Asun y Piedad.

Acknowledgements

I finally set out to write one of the most important (albeit the least scientific) sections of this thesis. Pursuing a PhD has been an intense learning journey, at many different levels, and I would not have made it this far without the support of a lot of people who I really want to thank.

The following body of work would not have been possible without the guidance and dedication of my thesis mentor, Paco Real. I want to thank you for giving me the opportunity of joining your laboratory. Your passion for research has been truly inspirational. Your energy, determination, and work-ethic have been stimulating, and have been essential in helping me become a better scientist.

I would like to acknowledge the members of my PhD thesis committee, Dr. Manuel Serrano, Dr. Marcos Malumbres, and Dr. Alberto Muñoz, as well as Dr. Ana Losada. Your feedback was paramount for the progress of my projects, and I want to thank you for all the knowledge and fresh ideas you brought into our meetings.

I also want to thank past and present members of the Epithelial Carcinogenesis Group for creating a warm, collaborative, and encouraging work atmosphere: Marinela and Julie (thanks for making my transition into the lab smoother), Marta (thanks for being such a good friend both in and out of the lab), Jarek (Montoya-evenings and guitar parties are not the same without you), Victor (you are always there to listen and give good advice), Miriam (you keep such a priceless attitude in science and in life), Laia (your talent and hard work are contagious), Eleonora (both the bladder team and lab outings have gained so much since you joined the lab), Ana (thanks for all your help and for being such a good bench and computer neighbor), Enrique (you got me to learn more bioinformatics that I ever thought I was capable of!), Luis (thanks for the endless jokes and interesting discussions), Ariel (for all your help and good humor), Xavi and Carme (for teaching me so many technical tricks and for your great attitude at all times), Natalia (thanks for always having an answer to my questions), Yoli (you were always so cheerful even though –or just because- I didn't work with mice), José María (thanks for always being so positive), Paola (for your wise advice and suggestions), Francesc (for giving the lab a musical touch), Isidoro (always ready to crack a joke), and adopted lab members Lina and Espe (thanks for sharing your expertise and being so easy-going). To everybody else who has been part of the team at some point (Andreia, Lorenzo, Raquel, Christie, Ainhoa, Thalía, Inma, María, Blanca, Ilaria, Ulisses…), thanks as well for your support.

I also want to put out a word of gratitude to the entire Melanoma group, for sharing reagents, giving great input during joint meetings, and because you are the best lab neighbors one could wish for. Thanks!

I would like to thank Dr. Núria Malats, as well as her group members Alessandra and Mirari, for working with Paco, Francesc, Ana, Inma and I in TERT-Escuela, such an innovative project mixing research and education.

There are a lot of people at the CNIO who have made my days here easier: Thank you, La Caixas 2010, for being such a cool group of people and making moving to Madrid and starting a PhD so much easier. Thanks, Rocío, for all your technical help with "chromosome counting" and for your accommodative and kind attitude. Thanks, Manu, Diego, and Chimo, for your help with confocal microscopy and the OPERA equipment. Lola, Ultan, thank you for your support with flow cytometry. Thanks as well to the Cytogenetics group, who are the best people to share our narrow corridor in tissue culture with! Thanks to everybody who ever shared a snack, a reagent, an early morning in tissue culture or a late night at the library. I don't want to forget members of the maintenance team, who couldn't be more efficient, as well as the cleaning staff, especially Gloria, who always has a kind word for everybody.

I wouldn't have embarked in the adventure of a PhD without the influence of former mentors and teachers: Gracias, Ana, porque las disecciones que hicimos contigo en el cole me despertaron la curiosidad sobre cómo funcionamos los seres vivos. Gracias, Ángel, por esos problemas de genética con orejas de elfo que me animaron a centrarme en lo molecular. Thanks, Wendy, for taking the risk and bringing Drosophila into the high-school lab: you got me hooked with that "lab assistant" role. Finally, thank you, Alison, for being a great core lab teacher, and even a better senior thesis advisor, and for introducing me to the wonderful world of yeast!

Quiero agradecer también a la gente maravillosa que he conocido durante estos años, que desarrollan su trabajo en el espacio de encuentro entre investigadores y público general: Víctor, Chema, Chechu, sois grandes profesionales y mejores divulgadores. Ojalá llegue a transmitir la mitad de pasión por la ciencia que vosotros. Ángela, Raúl, Almu, Dani, Alba, Alberto, Lucía, Vio, y todos los voluntarios de Escuelab: gracias por hacer posible este proyecto tan ilusionante de llevar la ciencia a las aulas.

A la gente de Madrid, muchas gracias por acogerme como una gata más. Quiero agradecer todo vuestro apoyo y los buenos momentos compartidos a mis queridos miembros de la Croquipandi (¡oé!): Baird (nuestro sempiterno líder aún en la distancia), Cruz, Moni, Mara y Nieves (y Albert, y Arduro), Javi, David, Vicen, Oscar (¡de marketing!), el otro Javi, Johny, Morgan, Pedro, Marta y los enanos. A mi gente de Colegios (del Mundo Unido), en especial a Olaya, Aitor, Berta, Andrew, Cris Santamaría y a mi Junta preferida (Jorge, Laura, Isaac, JJ): siempre es un placer trabajar con vosotros en el proceso de selección y en las actividades de la asociación y, por supuesto, también disfrutar juntos de los momentos más ociosos. A mis amigos del barrio, Ale, Bego, Vir, Berni, Ire, David, Muski, Silvi y Patri, gracias por hacerme sentir como en casa y por tantas risas compartidas.

También quiero agradecer su presencia constante a todos mis amigos de Princeton y de LPC: aunque estemos repartidos por el mundo y nos veamos tan poco, siempre os siento cerca; vuestro apoyo durante el doctorado ha sido muy importante. A mis amigos de Burgos, los de siempre (Nut, Jesús, Bego y Juanjo), gracias por estar siempre ahí, por escuchar mis historietas del labo, mis frustraciones con los experimentos y mis incertidumbres vitales, por las pelis y las cervezas. Os quiero.

Lo más importante, para el final: muchísimas gracias, familia. Gracias a mis abuelos, tíos y primos, por vuestro apoyo, tácito o explícito, y por alegraros genuinamente de mis logros. Gracias, Chava, por guardar ese examen de química orgánica. Gracias, Vio, por las conversaciones frikis durante las comidas, por los viajes, por las confidencias, por las pelis, por todo. Te quiero, hermanita. Papá, mamá, muchas gracias por cuidar tan bien de nosotras, por respetar mis decisiones y apoyarme siempre, por hacer todo lo posible por que Vio y yo seamos felices. Os quiero.

Gracias, Andrés, por aparecer en mi vida cuando más te necesitaba, y llenarla de azul y verde. Por ese punto de tranquilidad que le pones a todo, por tu confianza incondicional y tu optimismo inagotable. Gracias por hacerme reír, por los viajes, por los planes y, sobre todo, por llenarme de ilusión por el futuro pese a la incertidumbre. Te quiero.

¡MUCHÍSIMAS GRACIAS A TODOS POR ACOMPAÑARME EN ESTA ETAPA!

Table of contents

Summary

Despite the high incidence and economic cost of UBC, the molecular landscape of this tumor type remains relatively understudied. In this dissertation I describe the identification of novel genes involved in the development of UBC, using whole exome sequencing and a combination of targeted resequencing strategies. Our findings confirm previous studies pinpointing the chromatin remodeling pathway as frequently altered in UBC, including mutations in *ARID1A*, *KDM6A*, *CREBBP*, *EP300*, *MLL*, and *MLL3*. Moreover, we identified mutations in previously unreported genes belonging to this pathway (*MLL2*, *ASXL2*, and *BPTF*). Additionally, for the first time we found frequent mutations in DNA repair genes (*ATM*, *ERCC2*, and *FANCA*), as well as in subunits of the cohesin complex (*STAG1*, *STAG2*, *SMC1A*, and *SMC1*β). We did not identify significant differences in number or type of mutations according to tumor aggressiveness, patient age or smoking status. We analyzed mutations and loss of ARID1A expression to place them in the context of current bladder cancer molecular knowledge, finding they preferentially associated with the aggressive pathway of genetic progression. Regarding cohesin, we assessed the expression of its components in a cohort of 91 UBCs, finding low-frequency losses of SMC1, SMC3, RAD21, and PDS5B, as well as much more frequent losses of STAG2 expression. Intriguingly, we found that meiotic cohesin components SMC1β, REC8, and STAG3 are also expressed in UBC. We then focused on STAG2: mutations, predominantly truncating, were distributed all along the gene with a pattern characteristic of tumors suppressors, and were more frequent in non-aggressive UBCs. We also found evidence of genomic losses as a cause of STAG2 inactivation in a small subset of tumors. Loss of STAG2 was more common in non-aggressive tumors and associated with low proliferative index, mutations in *FGFR3*, and normal p53 expression. Importantly, STAG2 negative tumors frequently retained STAG1 expression, suggesting the competence of the cohesin complex. Contrary to what has been reported in other tumor types, STAG2 loss was not associated with aneuploidy in UBC tumors. Moreover, downregulating STAG2 *in vitro* did not cause significant changes in chromosome number or centromeric cohesion defects. As expected for a tumor suppressor, STAG2 overexpression impaired colony formation *in vitro*. Surprisingly, the same effect was seen upon STAG2 knockdown. RNA-sequencing of STAG2 knockdown cells did not reveal consistent transcriptional changes. Lastly, we found that loss of STAG2 associated with improved outcome in both NMIUBC and in MIUBC. However, it was an independent predictor of outcome only in the latter. Altogether, this work supports an important role of chromatin remodeling and cohesin components in UBC and contributes to refine the genetic events involved in UBC development/progression. Further studies are needed to unveil the mechanisms through which *ARID1A* and *STAG2* alterations contribute to UBC and to determine whether they can improve the molecular taxonomy of this tumor and can be used in patient management.

Resumen

Pese a la alta incidencia y coste económico del cáncer de vejiga (CV), el panorama molecular de este tipo de tumor permanece relativamente poco estudiado. En este trabajo describo la identificación de nuevos genes involucrados en el desarrollo de CV usando secuenciación de exomas y una combinación de estrategias de resecuenciación dirigida. Nuestros hallazgos confirman estudios previos que apuntaban a que las vías de remodelación de la cromatina sufre alteraciones frecuentes en CV, incluyendo mutaciones en *ARID1A*, *KDM6A*, *CREBBP*, *EP300*, *MLL*, y *MLL3* (Gui 2011). Además, hemos identificado alteraciones en genes no descritos que pertenecen a esta vía (*MLL2*, *ASXL2*, y *BPTF*). Por primera vez, hemos encontrado alteraciones frecuentes en genes involucrados en la reparación del ADN (*ATM*, *ERCC2*, y *FANCA*), así como en subunidades del complejo cohesina (*STAG1*, *STAG2*, *SMC1A*, y *SMC1*β). No hemos identificado diferencias significativas en el número o tipo de alteraciones de acuerdo a la agresividad del tumor o a la edad o exposición al tabaco del paciente. Hemos analizado más a fondo las mutaciones y pérdida de expresión de STAG2 para colocarlas en el contexto de las vías moleculares del CV conocidas, encontrando que se asocian preferentemente con la vía más agresiva de progresión genética. Para caracterizar en detalle el papel del complejo cohesina en CV, valoramos la expresión de los componentes del complejo en una población de 91 CV, hallando pérdidas poco frecuentes de SMC1, SMC3, RAD21, y PDS5B, así como pérdidas mucho más frecuentes de la expresión de STAG2. Curiosamente, hemos observado que SMC1β, REC8 y STAG3, componentes de la cohesina meiótica, también se expresan en CV. Hemos analizado las alteraciones y pérdida de expresión de STAG2 en una población más numerosa de CV, hallando tanto mutaciones como pérdidas genómicas que son responsables de la pérdida de expresión de STAG2 en CV. La mayoría de las mutaciones en *STAG2* son truncantes, se distribuyen a lo largo de la secuencia de la proteína en un patrón característico de los supresores tumorales y son más frecuentes en los CV no agresivos. La pérdida de STAG2 es más común en los tumores no agresivos y se asocia a mutaciones en *FGFR3*, expresión normal de p53 y un índice proliferativo bajo. Los tumores que no expresan STAG2 suelen retener la expresión de STAG1, lo que sugiere que en este contexto el complejo cohesina permanece activo. La pérdida de STAG2 no se correlaciona con la diferenciación de las células uroteliales. Al contrario de lo que se ha reseñado en otros tipos de tumores, la pérdida de STAG2 no causa aenuploidía o defectos en la cohesión centromérica en líneas celulares o tumores de CV. Como se esperaría de un supresor tumoral, la sobreexpresión de STAG2 impide la formación de colonias *in vitro*. Sorprendentemente, al silenciar la expresión de STAG2 se observa el mismo efecto. Sin embargo, la secuenciación de RNA de células en las que se había silenciado STAG2 no reveló cambios transcripcionales consistentes. Por último, hemos descrito que la pérdida de STAG2 también se asocia con un mejor resultado en el CV invasivo. En conjunto, este

trabajo apoya la importancia del papel de los remodeladores de cromatina y de los componentes de la cohesina en CV y contribuye a refinar los eventos genéticos involucrados en el desarrollo/progresión del CV. Hay que realizar más estudios para determinar si las alteraciones en *ARID1A* y *STAG2* contribuyen a la taxonomía molecular del CV y para mejorar el manejo de los pacientes.

Abbreviations

4EBP1	eIF4E-Binding Protein 1
ADC	Adencarcinoma
AKT	v-Akt murine thymoma viral oncogene homolog
ANOVA	Analysis of variance
APOBEC	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
APOBEC3B	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3B
ARID1A	AT rich interactive domain 1A
ARID1B	AT rich interactive domain 1B
ASXL2	Additional sex combs like 2
ATM	Ataxia telangiectasia mutated
ATP	Adenosine triphosphate
AURKA	Aurora kinase A
BAC	Bacterial artificial chromosome
BAF	BRG1-Associated Factor
BAIAP2L1	BAI1-associated protein 2-like 1
B-CAT	Branched chain amino-acid transaminase
BCG	Bacillus Calmette-Guerin
BNC2	Basonuclin 2
BPTF	Bromodomain PHD finger transcription factor
BRG1	BRM/SWI2-Related Gene 1
BRM	Brahma
BSA	Bovine serum albumin
CCDC88B	Coiled-coil domain containing 88B
CCND1	Cyclin D1
CD83	Cluster of differentiation 83
CDK	Cyclin-dependent kinase
CDK2	Cyclin-dependent kinase 2
CDK4	Cyclin-dependent kinase 4
CDK6	Cyclin-dependent kinase 6
CDKi	Cyclin-dependent kinase inhibitor
CDKN2A	Cyclin-dependent kinase inhibitor 2A
CDKN2B	Cyclin-dependent kinase inhibitor 2B
cDNA	Complementary DNA
CGARN	The Cancer Genome Atlas Research Network
CGH	Comparative genomic hybridization
CI	Confidence intervals
CIS	Carcinoma in situ
c-Myc	Cellular myelocytomatosis oncogene
CNIO	Centro Nacional de Investigaciones Oncológicas
COL1A1	Collagen, type I, alpha 1
CREBBP	CREB binding protein
CTCF	CCCTC-binding factor
DAB	3,3'-Diaminobenzidine
DAPI	4,6-diamidino-2-phenylindole
DAPK1	Death-associated protein kinase 1

DBC1	Deleted in bladder cancer protein 1
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
E2F	E2 transcription factor
E2F3	E2F transcription factor 3
ECAD	E-cadherin
EDTA	Ethylenediaminetetra-acetic acid
EGFR1	Epidermal growth factor receptor 1
EP300	E1A binding protein p300
EPICURO	Spanish Bladder Cancer Study
ERBB2	v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2
ERCC2	Excision repair cross-complementation group 2
ERK	Extracellular signal-regulated kinase
ERK1/2	Extracellular signal-regulated kinase 1/2
ESCO1	Establishment of sister chromatid cohesion N-acetyltransferase 1
ESCO2	Establishment of sister chromatid cohesion N-acetyltransferase 2
ESPL1	Extra spindle pole bodies homolog 1
FANCA	Fanconi anemia, complementation group A
FBS	Fetal Bovine Serum
FDR	False discovery rate
FGFR	Fibroblast growth factor receptor
FPKM	Fragments per kilobase of exon per million fragments mapped
FRP1	Free running period 1
FYN	Tyrosine-protein kinase Fyn
G	Grade
G1	Gap 1 phase of the cell cycle
GEO	Gene Expression Omnibus
GLI2	GLI family zinc finger 2
GO	Gene Ontology
GSTM1	Glutathione S-transferase mu 1
GTP	Guanosine 5'-triphosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPRT	Hypoxanthine phosphoribosyltransferase 1
HPV	Human papiloma virus
HRAS	v-Ha-ras harvey rat sarcoma viral oncogene homolog
HRP	Horseradish peroxidase
IFI6	Interferon, alpha-inducible protein 6
IGF1R	Insulin-like growth factor 1 receptor
IHC	Immunohistochemistry
INI1	Integrase Interactor 1
ISBLAC	Integrated Study on Bladder Cancer
KDM6	Lysine (K)-specific demethylase 6
KEGG	Kyoto Encyclopedia of Genes and Genomes
Ki67	Marker of proliferation Ki-67

KO	Knock out
KRAS	Kirsten rat sarcoma viral oncogene homolog
KRT10	Keratin 10
KRT14	Keratin 14
KRT18	Keratin 18
KRT20	Keratin 20
KRT5	Keratin 5
KW	Kruskal Wallis test
LOH	Loss of heterozygosity
MAPK	Mitogen-activated protein kinase
MAPK15	Mitogen-activated protein kinase 15
MAU2	MAU2 chromatid cohesion factor homolog
MEF	Mouse embrionic fibroblast
MET	Met proto-oncogene
MFGE8	Milk fat globule-EGF factor 8 protein
MI	Muscle-invasive
MIUBC	Muscle-invasive urothelial bladder cancer
MLL	Myeloid/lymphoid or mixed-lineage leukemia
MLL2	Myeloid/lymphoid or mixed-lineage leukemia 2
MLL3	Myeloid/lymphoid or mixed-lineage leukemia 3
mRNA	Messenger RNA
MTMR3	Myotubularin related protein 3
mTOR	Mammalian target of rapamycin
MVAC	Mehtrotrexate, vinblastine, doxorubicin, and cisplatin
MW	Mann-Whitney U test
NAT2	N-acetyltransferase 2
NFDM	Non-fat dried milk
NIPBL	Nipped-B homolog
NMI	Non muscle-invasive
NMIUBC	Non muscle-invasive urothelial bladder cancer
NPR2	Natriuretic peptide receptor 2
NRAS	Neuroblastoma RAS viral (v-ras) oncogene homolog
NS:S	nonsynonymous:synonymous ratio
NU	Normal urothelium
OR	Odds ratio
P	Probability values
p110	Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha protein
p16	Cyclin-dependent kinase inhibitor 2A protein
p21	Cyclin-dependent kinase inhibitor 1 protein
P33ING1	p33 inhibiton of growth 1
p53	Tumor protein 53
p63	Tumor protein 63
p85	Phosphatidylinositol 3-kinase regulatory subunit protein
PARP	Poly (ADP-ribose) polymerase
PARP1	Poly (ADP-ribose) polymerase 1

PBAF	SWI/SNF-B chromatin-remodeling complex
PBS	Phosphate-Buffered Saline
PCA	Principal Component Analysis
P-CAD	Cadherin 3, type 1
PCR	Polymerase chain reaction
PDK1	Pyruvate dehydrogenase kinase, isozyme 1
PDS5A	PDS5, regulator of cohesion maintenance, homolog A
PDS5B	PDS5, regulator of cohesion maintenance, homolog B
PI3K	Phosphoinositide-3 kinase
PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha
PIM1	Proto-oncogene serine/threonine-protein kinase Pim-1
PIP2	Phosphatidylinositol 4,5-bisphosphate
PIP3	Phosphatidylinositol (3,4,5)-trisphosphate
PKC	Protein kinase C
PLK1	Polo-like kinase 1
PP2A	Protein phosphatase 2A
pRB	Retinoblastoma protein
PTCH	Patched
PTEN	Phosphatase and tensin homolog
PTPRD	Protein tyrosine phosphatase, receptor type, D
RAD21	RAD21 homolog
RAD51	RAD51 recombinase
RB	Retinoblastoma
RB1	Retinoblastoma 1
REC8	REC8 meiotic recombination protein
RHOBTB3	Rho-related BTB domain containing 3
RNA	Ribonucleic acid
RT-qPCR	Real-time reverse transcription polymerase chain reaction
S	Synthesis phase of the cell cycle
S6K	S6 kinase
SCC	Squamous cell carcinoma
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SGO1	Shugoshin 1
shNT	Non-targeting shRNA control
shRNA	Short hairpin RNA
SLC14A	Solute carrier family 14 A
SMC1A	Structural maintenance of chromosomes 1A
SMC1β	Structural maintenance of chromosomes 1β
SMC3	Structural maintenance of chromosomes 3
SNF5	Sucrose nonfermenting protein 5
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
STAG1	Stromal antigen 1
STAG2	Stromal antigen 2
STAG3	Stromal antigen 3

SWI/SNF	Switch/Sucrose NonFermentable
T	Stage
TACC3	Transforming, acidic coiled-coil containing protein 3
TBS-T	Tris buffered-saline - tween
TERT	Telomerase reverse transcriptase
TFEB	Transcription factor EB
TMA	Tissue microarray
TOP2A	Topoisomerase II alpha
TP53	Tumor protein 53
TRIO	Trio Rho guanine nucleotide exchange factor
TSC1	Tuberous sclerosis 1
TSC22D2	TSC22 domain family, member 2
TSC22D3	TSC22 domain family, member 3
TUR	Transurethral resection
UBC	Urothelial bladder cancer
UCSF	University of California San Francisco
UPK3B	Uroplakin 3B
US	United States
WAPL	Wings apart-like homolog
WES	Whole Exome Sequencing
WHO	World Health Organization
WNT6	Wngless-type MMTV integration site family, member 6
WNT7A	Wngless-type MMTV integration site family, member 7A
WNT9A	Wngless-type MMTV integration site family, member 9A
WT	Wild type

Introduction

1. URINARY BLADDER BIOLOGY

The urinary bladder, located above and behind the pelvic pubic bone, is a distensible organ responsible for the collection and storage of the urine produced at the kidneys, as well as its subsequent disposal through the urethra. Bladder architecture is adapted to allow the fulfillment of these functions, with a hollow space surrounded by a layer of muscle. It is organized in four distinct layers: the urothelium, the lamina propria, the detrusor muscle, and the perivesical fat (Figure 1a).

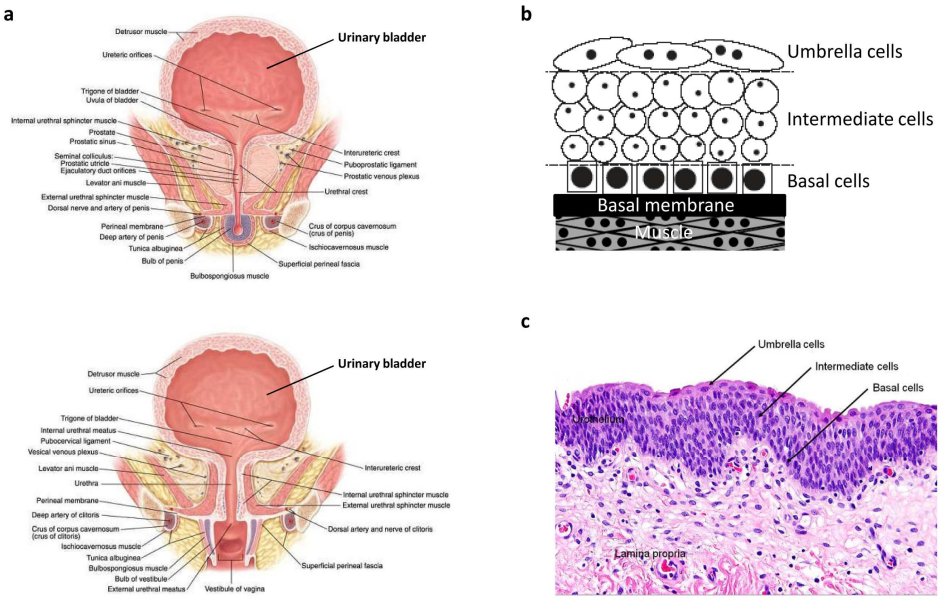

Figure 1. Anatomy and histology of the urinary bladder. (a) Anterior anatomy of the male (top) and female (bottom) urinary bladder. Adapted from Tank et al. 2009. (b) Schematic representation of the urothelium. Modified from Laguna et al. 2006. (c) Hematoxylin-eosin staining of normal urothelium. Adapted from Krueger et al. 2012.

1.1. The urothelium

The urothelium, a multilayered epithelium, is comprised of 3-6 cell layers depending on the degree of distension of the organ (Jost et al. 1989). It lines the whole urogenital tract, starting at the renal pelvis and reaching as far as the urethra. There are three types of cells in the urothelium (Figure 1b, c). The most superficial and differentiated ones are called umbrella cells, which have a diameter of 50-120 µm, are often multinucleate, form a single layer, and are characterized by the expression of four highly-conserved uroplakin proteins and low molecular weight cytokeratins (CK18, CK20) (Moll et al. 1990; Schaafsma et al. 1990; Apodaca 2004). Each umbrella cell covers several underlying intermediate cells, which measure 20 µm across. Lastly, basal cells, which form a single layer, have a diameter of 5-10 µm (Lewis 2000). Both basal and intermediate cells express

p63 and high molecular weight cytokeratins (KRT5, KRT10, KRT14) (Moll et al. 1982; Yang et al. 1998; Castillo-Martin et al. 2010). Originally, morphological observations and labeling experiments suggested that basal cells migrate towards the bladder lumen and fuse to give rise to multinucleated umbrella cells (Martin 1972). However, more recent studies suggest that umbrella and basal cells have different origins (Signoretti et al. 2005; Dancik et al. 2013).

1.2. Other bladder structures

Beneath the urothelium lies the lamina propria, which is composed of stromal connective tissue of varying densities, as well as some discontinuous stretches of smooth muscle usually associated with blood vessels. Underneath the lamina propria lies the detrusor muscle of the bladder, which combines circular and longitudinal layers of smooth muscle bundles without concrete orientation. Finally, there is a layer of adipocytes, fibrous tissue, and blood vessels surrounding the bladder that are known as the perivesical tissue.

2. BLADDER CANCER INCIDENCE AND MORTALITY

Bladder cancer is the second most frequent urogenital tract tumor in men (Sieguel et al. 2014). It is estimated that over 386,000 patients are diagnosed and 150,000 die from bladder tumors every year worldwide (Jemal et al. 2011). The difference between incidence and mortality is indicative of a prolonged progression interval. Indeed, the average 5-year survival rate in Europe is around 70% (Sant et al. 2003).

The incidence of bladder cancer is higher in males than females: the 2012 age-standardized incidence estimated rates per 100,000 European individuals were 26.9 in males and 5.3 in females (Ferlay et al. 2013). The areas with the highest incidences are North Africa, Europe and North America (Jemal et al. 2011). However, the likelihood of detecting low-grade tumors varies across world regions, so studying global mortality rather than diagnosis rates is usually more informative. Examining worldwide male mortality rates, the highest values are found in Egypt (16.3 per 100,000), followed by several European countries (Jemal et al. 2011). Within Europe, Spain has one of the highest male incidence rates: 32.71 cases per 100,000 inhabitants, as opposed to the European average of 26.9 (Miñana et al. 2014).

Interestingly, even though women are less likely to develop bladder tumors, they exhibit higher UBC-related mortality rates than men even after adjusting for tumor stage and grade (Cook et al. 2011). A later diagnosis, physiological differences in bladder muscle strength, and other sociological, hormonal and environmental factors have been suggested to play a role (Fajkovic et al. 2011).

3. BLADDER CANCER ETIOLOGY

About half of all bladder cancer cases in Africa and the Middle East can be attributed to chronic *Schistosoma haematobium* infection (Parkin 2006). Recurrent *Schistosoma* infection causes chronic cystitis, favoring the development of bladder tumors. Tobacco smoking, exposure to high arsenic concentrations in drinking water, and aromatic amine work-related exposure are the major risk factors in Western countries (Letašiová et al. 2012). Smoking accounts for 50% of bladder cancer cases in men (Zeegers et al. 2000), while occupational exposures are responsible for fewer than 10% of the cases in Western Europe males (Kogevinas et al. 2003).

There is a positive correlation between daily number of cigarettes smoked or number of smoking years and the risk of developing bladder cancer (Silverman et al. 2006). Although the risk is significantly reduced as soon as one year after quitting, it does not reach the levels of individuals who have never smoked; only a 60% reduction in risk is observed even 25 years after quitting (Brennan et al. 2000). Unfiltered cigarettes pose about twice as much risk as filtered ones (Wynder et al. 1988) and so does black as opposed to blond tobacco (De Stefani et al. 1991). A more recent study in Spain found that although risk is 40% higher for smokers of black tobacco than for those using blonde varieties, this tendency is not significant (Samanic et al. 2006).

Aromatic amines have been shown to play a major role in the development of UBC (reviewed in Vineis and Pirastu 1997). Aromatic amines are present at low levels in tobacco smoke and are transformed into metabolic intermediates that form DNA adducts (Talaska et al. 1991). Since aromatic amines are metabolized by the kidney, they become potential carcinogens in the upper and lower urinary tracts (reviewed in Jung and Messing 2000). Occupational exposure to aromatic amines, polycyclic aromatic or chlorinated hydrocarbons, mainly affect workers in the paint, dye, metal and petroleum processing jobs (reviewed in Burger et al. 2013). Other less characterized factors that might affect the risk of developing UBC include diet and fluid intake. The intake of liquids might decrease said risk by further diluting carcinogens in urine but chlorinated water negatively affects such risk and both factors have never been studied simultaneously. On the other hand, the presence of arsenic in drinking water has been proven to cause UBC in regions subjected to elevated exposure, such as Bangladesh and Chile (Fernandez et al. 2012).

There are also genetic factors associated with the development of UBC. Highly prevalent polymorphisms decreasing NAT2 and GSTM1 activity have been associated to a ca. 30% increase in risk of developing bladder cancer (García-Closas et al. 2005). NAT2 and GSTM1 detoxify aromatic amines, decreasing carcinogenic DNA adduct formation. Additionally, recent genome-

wide association studies have identified other sequence variants linked to genetic predisposition to developing UBC (Kiemeney et al. 2008, 2010; Rafnar et al. 2009; Wu et al. 2009; Rothman et al. 2010; Garcia-Closas et al. 2011). One of the responsible genes accounting for such susceptibility was subsequently identified: *SLC14A*, which encodes a urea transporter that controls urine concentration, affecting bladder exposure to carcinogens (Rafnar et al. 2011).

4. BLADDER TUMOR CLASSIFICATION

The majority of bladder tumors arise in the urothelium, being designated as urothelial bladder cancer (UBC). Bladder squamous cell carcinomas (SCC) and adenocarcinomas (ADC) are less frequent tumors with features of squamous epithelial cells and bladder mucosa lining cells respectively. Most bladder tumors associated with tobacco are UBCs although urothelial tumors can present focally variable degrees of squamous differentiation; the majority of worldwide cases related to *Schistosoma* infection are SCCs (Sliverman et al. 2006). This is why in Egypt 75% of bladder tumors are SCCs. In the Western world, more than 90% of newly diagnosed cases are UBCs and pure SCCs represent 5-7% of the cases. Bladder ADCs are the least frequent tumors, representing 2% of all cases (Silverman et al. 2006). Other extremely rare non-urothelial bladder carcinomas include small cell carcinomas, lymphomas, and sarcomas. Morphologically, three categories of UBCs can be distinguished. Papillary tumors are the most frequent type; they develop projections towards the bladder lumen and tend to grow slowly. Solid tumors, which are less common but more aggressive, tend to grow into deeper bladder layers. The last type are in situ carcinomas (CIS) which are flat and only involve the most superficial fraction of the urothelium, but are nonetheless very aggressive.

4.1. Staging and grading of bladder tumors

The classification of bladder tumors is somewhat controversial and has been revised several times over the last 25 years by pathologists, clinicians and researchers. Usually, the main factors taken into account for bladder tumor categorization are stage and grade. Staging of UBCs is performed according to the degree of tissue invasion, whereas grading takes into consideration the morphology and degree of differentiation of tumor cells.

The following stages can be distinguished in UBC (reviewed in Edge et al. 2010, see Fig. 2):

-Tis: This stage corresponds to CIS, flat carcinomas restricted to the urothelium.
-Ta: These are papillary tumors restricted to the most superficial layer of the bladder lining.
-T1: These solid tumors have surpassed the lamina propria, invading the connective tissue beneath the urothelium.

-T2: The solid tumor has reached and grown into the surface of the bladder muscle.
-T3: This stage designates solid tumors that have invaded the bladder perivesical fat.
-T4: These solid tumors have grown until reaching neighboring organs within the pelvis, such as the prostate or the uterus.

According to this classification, we can classify tumors as non-muscle invasive (NMI), including Tis, Ta, and T1, and muscle invasive (MI) comprising T2, T3 and T4 (Cheng et al. 2009).

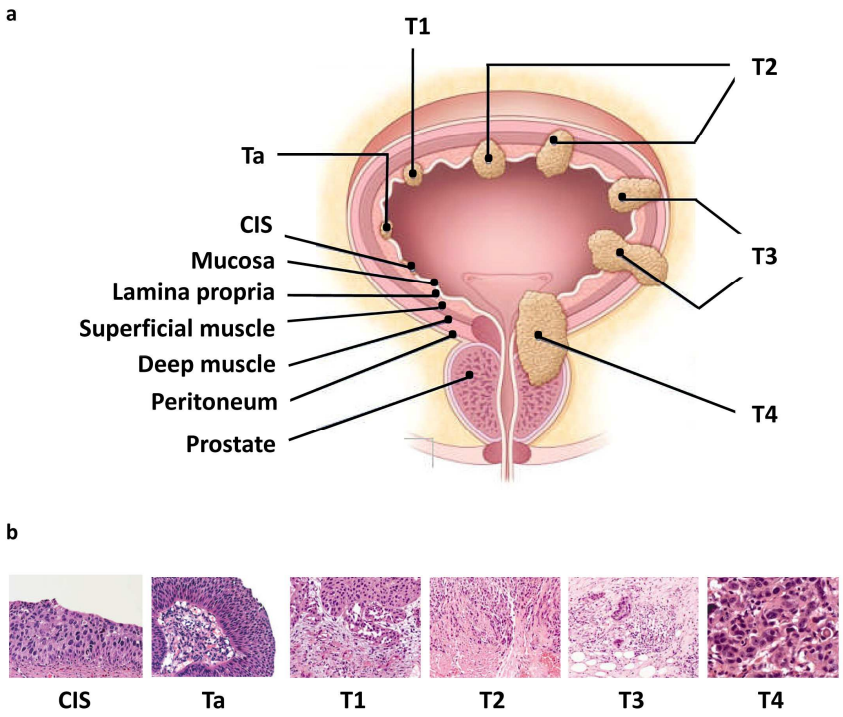

Figure 2. Staging of UBC. (a) Diagramatic representation of UBC staging, modified from Action on Bladder Cancer (www.actiononbladdercancer.org). (b) Hematoxylin-eosin staining of UBC of different stages. CIS and Ta adapted from Reis 2011, T1-T4 modified from Cheng et al. 2009.

Over time, many different grading classifications have been developed, differing mainly in the nomenclature ascribed to preneoplastic lesions and the number of categories used to sort UBCs.

1973 WHO grading
Urothelial papilloma:
Grade 1: well differentiated
Grade 2: moderately differentiated
Grade 3: poorly differentiated
2004 WHO grading
Flat lesions:
Hyperplasia (flat lesion without atypia or papillary)
Reactive atypia (flat lesion with atypia)
Atypia of unknown significance
Urothelial dysplasia
Urothelial CIS
Papillary lesions:
Urothelial papilloma (which is a completely benign lesion)
Papillary urothelial neoplasm of low malignant potential
Low-grade papillary urothelial carcinoma
High-grade papillary urothelial carcinoma
WHO = World Health Organisation; CIS = carcinoma in situ.

Table 1. 1973 and 2004 WHO Classifications of UBC. Extracted from Babjuk et al. 2011.

The most recent grading is the 2004 World Health Organization (WHO) classification, which differentiates between NMI and MI urothelial tumors (Table 1) (Montironi and López-Beltrán 2005). However, it has been recommended to use it together with the 1973 WHO classification until it is clinically validated (Montironi and López-Beltrán 2005). The 1973 WHO classification uses three grades: G1 and G2 are generally considered low grade, and refer to cells

that are well differentiated and exhibit few mitoses, and G3 are of high grade with poorly differentiated cells harboring aberrant nuclei and frequent mitoses (Mostofi et al. 1973).

Most clinicians and researchers use a combination of grading and staging to describe the cases they study according to how life-threatening the tumors are: low risk (LR) (TaG1, TaG2, and T1G2) and high risk (HR) (TaG3, T1G3, and all ≥T2 tumors regardless of grade). Throughout this thesis, the terms non-aggressive (tumor of Ta stage and low grade) and aggressive (tumor of high grade regardless of stage) will also be used.

5. BLADDER CANCER DIAGNOSIS AND CLINICAL MANAGEMENT

The most common symptom of bladder cancer is painless hematuria. Approximately 20% of patients with gross hematuria and 5% of those with microscopic hematuria have a bladder tumor (reviewed in Chou and Dana 2010). At diagnosis, 75-85% of UBC cases present NMI tumors, of which 31-78% will progress and 1-45% will recur within 5 years from initial diagnosis (Babjuk et al. 2011; Sylvester et al. 2006). Approximately half of MI tumors metastasize, frequently leading to death. Upon detection, bladder tumors are usually removed using transurethral resection (TUR), and diagnostic biopsies are taken if CIS is suspected (Kamat et al. 2013).

5.1. Management of NMIUBC

In NMIUBC, current clinical practice advises that a complete TUR is performed in patients with papillary tumors and, if the tumor is of T1 stage and/or high grade, a second TUR is recommended 2-6 weeks after the initial one (Babjuk et al. 2011). This is due to the fact that these tumors might present several locations, resulting in TUR leaving behind residual cancerous cells. It is recommended to administer one dose of intravesical chemotherapy immediately after resection. The most commonly used drug for this purpose is mitomycin-C, although epirubicin or doxorubicin are also frequently administered with comparable results (Babjuk et al. 2011). This should be followed by either further chemotherapy instillations (to prevent recurrence) or at least 12 months of bacillus Calmette-Guerin (BCG) intravesical therapy if there is intermediate or high risk of progression (Babjuk et al. 2011). The same treatment is suggested in the presence of CIS.

BCG attaches to integrin and fibronectin receptors in bladder tumoral cells, stimulating local cytokine release that triggers non-specific defense mechanisms. This results in the elimination of malignant cells and thus a delay or even completely prevention of tumor progression (Babjuk et al. 2011). BCG induction is performed for 6 consecutive weeks and periodical maintenance instillations lasting from 1-3 years are recommended (Babjuk et al. 2011). BCG instillation causes

frequent side effects that often lead to treatment withholding during the 6 initial months; the main serious complications are cystitis and sepsis and occur in about 5% of the cases (van der Meijden et al. 2003). Therefore, each patient should be carefully assessed for concurrent pathologies (such as post-TUR infections), recurrence and progression risks before prescribing BCG instillations.

If BCG treatment fails in patients with NMIUBC who present a very high risk of progression, cystectomy might be considered (Babjuk et al. 2011). This procedure has important morbidities: patients develop urinary tract infection or septicemia that required hospitalization in 30% of cases (van Hemelrijck et al. 2013), so it should only be prescribed to patients who would truly benefit from it. Factors leading to a recommended cystectomy in patients with T1 tumors include the presence of multiple or extensive tumor foci that prevent complete resection, recurrent disease after intravesical therapy or upon a second TUR, deep penetration into the lamina propria, invasion of the lymphatic space or the prostate, development of CIS, and patient failure to adequately follow the prescribed treatment (Shariat 2007).

About 20% of patients with NMI tumors are cured after the first TUR, but 50-60% undergo recurrence and 10-15% progress to develop MI tumors, which results in death in 50% of the cases (Sylvester et al. 2006). Such a heterogeneous prognostic landscape translates into a need for lifelong surveillance of the patients. This, together with potential complications, raises the costs derived from bladder cancer (reviewed in Svatek et al. 2014). A study performed in the United States demonstrated that the medical care of a bladder cancer patient can cost between $US 65,158 and $US 120,684 (Avritscher et al. 2006). The most cost-effective technique to assess recurrence is cystoscopy, which allows easy determination of number, location, and size of tumors, but also entails significant discomfort for the patients (Kamat et al. 2013). A downside of this technique lies in its failure to detect CIS, which has an appearance that is very similar to that of normal urothelium (Witjes 2004). To detect this type of tumor, photosensitising agents such as 5-aminolevulinic acid or hexyl aminolevulinate are used in combination with fluorescence microscopy (reviewed in Rink et al. 2013). These drugs work by stimulating tumor-cell selective accumulation of porphyrins, which emit in red wavelengths upon excitation with blue light (Kausch et al. 2009). This technique raises the cost of the procedure but it also results in fewer recurrences, with the associated savings, making it at least cost-neutral (Rink et al. 2013). There is still some debate as to whether photodynamic diagnostic cystoscopies should be the standard of care for NMIUBC patients.

5.2. Management of MIUBC

The treatment of MI tumors differs depending on whether the disease is localized or metastatic. For localized disease, radical cystectomy is the most common treatment (Stenzl et al. 2011). Whenever possible, a fraction of functioning urethra and its associated nerves should be preserved to allow for urinary diversion. The most common type of urinary diversion is the orthotopic neobladder substitution, which does not compromise patient prognosis and contributes to an improved life quality after cystectomy (Stenzl et al. 2010). Even though radiotherapy is sometimes administered before surgery, it has not been shown to increase survival and therefore is not currently advised for fit patients (Stenzl et al. 2011). However, a recent report argues that sensitizing tumors to radiation with fluorouracil or mitomycin C chemotherapy reduces the need for radical cystectomy, although their findings do not reach statistic significance (James et al. 2012). This strategy might prove adequate in cases where a cystectomy is not advised, such as elderly or frail patients.

There is some evidence that patient outcome can improve if radical cystectomy is preceeded by neoadjuvant containing chemotherapy, except for patients suffering from impaired kidney function or a weakened physical state (Stenzl et al. 2011). This reduces the risk of progression of undetectable micrometastasis in local MIUBC and increases the survival of patients with metastatic MIUBC. The first-line treatment for both patient groups is combination chemotherapy containing cisplatin. Cisplatin is administered together with gemcitabine or with methotrexate, vinblastine, and adriamycin (MVAC); the latter combination results in a slightly longer overall survival (15.2 months vs. 14.0 months) but the administration with gemcitabine is less toxic, so it is becoming the new treatment standard (von der Maase et al. 2005).

Lymphadenectomy should be performed for pelvic nodes upon cystectomy, both to optimize the staging of the tumor, and to improve survival (reviewed in Tilki et al. 2013). Some studies suggest that lymphadenectomy might improve survival regardless of tumor stage. However, the available evidence is mainly retrospective and of low quality, and there are no guidelines to standardize the procedure, so further trials are needed to determine the extent of the surgery and the optimal number of nodes to be removed.

When a patient with metastatic UBC is considered ineligible for cisplatin treatment, the combination of carboplatin and gemcitabine is advised. Upon progression after receiving first-line treatment, the recently approved second-line drug (vinfluvine) or inclusion in approved clinical trials for other drugs are recommended (Stenzl et al. 2011). Such clinical trials are starting to

include drugs targeting some of the underlying genetic alterations, such as those affecting *FGFR3* or *mTOR* (Bellmunt et al. 2013; Gust et al. 2013; Seront et al. 2013).

6. A MORPHOGENETIC VIEW OF UBC TUMOR PROGRESSION

The current paradigm is that there are two genetic pathways of tumor progression from normal urothelium (reviewed in Castillo-Martin et al. 2010; McConkey et al. 2010). In short, the non-aggressive pathway involves mainly oncogenes, with activating mutations in *FGFR3* and *PIK3CA* being the most common alterations. These mutations are associated with the development of tumors of low stage, which are usually papillary, and present a low likelihood of progression. A small proportion of them acquires genetic alterations such as *CDKN2A* losses and mutations in *TP53*, resulting in the development of invasive, high stage tumors. The aggressive pathway involves predominantly tumor suppressors: mutations in *TP53*, *PTEN*, and inactivation of the *RB* pathway are frequent. These tumors are aggressive, likely to progress, and are commonly detected after having invaded muscle, although it is thought that CIS can be a precursor of tumors belonging to this pathway. Both pathways share early losses of chromosome 9 and mutations in the *TERT* promoter. Additionally, there is extensive evidence proving that non-aggressive tumors are genomically stable, as opposed to aggressive tumors.

6.1. Alterations common to both pathways

The most common genomic alterations in all subtypes of UBC affect chromosome 9, which has been found to be totally or partially lost in tumors of all stages and grades (Spruck et al. 1994; Hartmann et al. 2002; Mhawech-Fauceglia et al. 2006). The frequently deleted regions include 9p22–23 in the short arm, and 9q11–13, 9q12–13, 9q21–22, and 9q34 in the long arm of the chromosome. They encompass at least 7 known tumor suppressor genes, namely *BNC2*, *DAPK1*, *DBC1*, *CDKN2A/B*, *PTPRD*, and *TSC1*, which could be responsible for the initial transformation of urothelial cells (Beothe et al. 2012). Consistently, bioinformatic analysis has suggested that loss of heterozygosity in chromosome 9 is one of the earliest genetic alterations in UBC pathogenesis (Bulashevska et al. 2004).

Mutations in the promoter of *TERT* have recently been reported to occur in 70% of UBC tumors (Kinde et al. 2013; Liu et al. 2013; Rachakonda et al. 2013; Allory et al. 2014; Hurst et al. 2014). A low proportion of control subjects without UBC may also present with *TERT* mutations. Altogether, these findings suggest that *TERT* promoter mutations are also a very early event in UBC shared by tumors belonging to both molecular pathways. Mutations have been detected in

exfoliated urine cells, making *TERT* promoter mutations a potential non-invasive diagnostic marker. In melanoma, mutations have been proposed to create novel transcription factor binding sites that could upregulate TERT expression (Horn et al. 2013; Huang et al. 2013) but no relationship was found between the presence of mutations and increased mRNA levels in UBC (Allory et al. 2014). Still, the functional significance of these mutations is not fully characterized.

6.2. The non-aggressive pathway

6.2.1. *FGFR3*

Fibroblast Growth Factor receptors (FGFRs) are tyrosine kinases consisting of three immunoglobulin-like domains, a transmembrane domain, and a cytoplasmic domain with tyrosine kinase activity (Schlessinger 2000). There are four FGFR family members, namely FGFR1-4. Ligand binding induces receptor dimerization, resulting in several phosphorylation events that lead to receptor activation (Haughsten et al. 2010) and downstream signaling through the FRS2 adaptor protein. FGFR3 is the most predominantly expressed member of the FGFR family in the urothelium. *FGFR3* is subjected to alternative splicing of its immunoglobulin-like domain III, producing the IIIb or IIIc isoforms, which have different tissue-specific expression and ligand specificities (reviewed in Knowles 2007). The IIIb isoform of *FGFR3* is expressed in epithelial cells.

Activating mutations of *FGFR3* in UBC were first identified in the late 90s (Cappellen et al. 1999). Between 60-65% of UBCs harbor activating, oncogenic point mutations in *FGFR3*. The absence of *FGFR3* mutations in normal urothelium from UBC patients suggests that they occur after chromosome 9 losses and *TERT* promoter mutations (Otto et al. 2009). *FGFR3* mutations are clustered in three hotspots and commonly involve the substitution of certain residues with Cys (van Rhijn et al. 2004; Kompier et al. 2011). In the extracellular domain, exon 7 harbors frequent R248C (6%) and S249C (44%) alterations. Exon 10, which maps to the transmembrane domain, commonly exhibits G372C (3%), Y375C (11%), and A393E (0.4%) substitutions. Finally, in the tyrosine kinase domain, exon 15 frequently presents K652M (0.8%) mutations. Additional alterations mapping to the same hotspots occur at lower frequencies (Knowles 2007). Mutations in the extracellular or transmembrane domains of FGFR3 (exons 7 and 10) often result in spontaneous dimerization, while mutations in the kinase domain (exon 15) lead to constitutive receptor activation (Cappellen et al. 1999; Adar et al. 2002). Mutations are associated with *FGFR3* mRNA and protein overexpression. (Tomlison et al. 2007). Signaling downstream of FGFR3 includes the PI3K/AKT, the RAS/MAPK, and the PKC pathways (Haughsten et al. 2010). Out of all

mutated tumors, 1/3 present concomitant mutations in other frequently altered genes in bladder cancer, while the remaining 2/3 harbor *FGFR3* mutations exclusively (Kompier et al. 2011).

Mutations in *FGFR3* are more common in non-aggressive tumors, particularly in those of lower stage and grade, as described in two large studies including cases representative of the full UBC spectrum (Hernandez et al. 2006; Kompier et al. 2011). Moreover, *FGFR3* alterations associate with a decreased risk of progression and longer survival (Hernandez et al. 2006; Kompier et al. 2011). Interestingly, overexpression of wild-type FGFR3 occurs in about 40% of tumors, usually associated with aggressive features (Tomlinson et al. 2007).

FGFR3 amplifications have been reported in 3% of UBCs (Fischbach et al. 2014). Additional *FGFR3*-related alterations in UBC involve gene fusions. *FGFR3-TACC3* and *FGFR3-BAIAP2L1* fusions have been identified in UBC cell lines, and 1/32 tumors without *FGFR3* mutations that over expressed the protein by IHC was shown to harbor a *FGFR3-TACC3* fusion (Williams et al. 2013). Subsequently, 2/42 tumors (1 NMI and 1 MI) have been show to harbor *FGFR3-TACC3* fusion transcripts (Guo et al. 2013). This results in increased expression of TACC3, a protein involved in spindle stabilization during chromosome segregation.

6.2.2. *PIK3CA*

The PI3K pathway is crucial for regulation of cell survival and growth (Engelman 2009). There are three classes of PI3K enzymes. Class I PI3Ks form heterodimers, which are composed of a catalytic and a regulatory subunit. Class IA PI3Ks are characterized by being activated by receptor tyrosine kinases (García et al. 2006). In Class IA PI3Ks, the regulatory subunit is p85, whereas the catalytic protein is p110 (Bader et al. 2005). p110-α, encoded by *PIK3CA*, has been found to be mutated in 13-25%% of UBC (López-Knowles et al. 2006; Platt et al. 2009; Kompier et al. 2011). The discrepancies among studies most likely are due to the different methods employed for mutation detection. The most commonly mutations are E542K, E545K (both in exon 9), and H1047R (exon 20). Exon 9 encodes the helicase domain, where 80% of the mutations occur in UBC, whereas exon 20 corresponds to the kinase domain. The effect of the mutations is an oncogenic increase in PI3K activity (Bader et al. 2006; Zhao and Vogt 2008). Additionally, several studies have reported amplifications of 3q, the chromosomal arm where *PIK3CA* localizes (Koo et al. 1999; Cazier et al. 2014). Upon activation by membrane receptors, class I PI3Ks catalyze the generation of phosphatidylinositol-3,4,5-triphosphate (PIP3) from phosphatidylinositol-4,5-bisphosphate (PIP2) (Bader et al. 2005). PIP3 binds to PDK1 and AKT, and AKT is activated by

PDK1-mediated phosphorylation. A phosphorylation cascade involving mTOR, S6K, and 4EBP1 results in the activation of protein translation (Bader et al. 2005).

Mutations in *PIK3CA* frequently co-occur with *FGFR3* mutations, suggesting that *PIK3CA* alterations are also an early event in UBC development (López-Knowles et al. 2006; Kompier et al. 2011). Additionally, this might imply that the oncogenic effect of mutations in *FGFR3* and *PIK3CA* cooperate in the non-aggressive pathway of UBC development. *PIK3CA* alterations are more common in low grade tumors, but they are not significant predictors of outcome among NMIUBC (Kompier et al. 2011).

6.2.3. *RAS* genes

Mutations in *HRAS, KRAS,* or *NRAS* occur with an overall 11% prevalence in UBC (Jebar et al. 2005; Kompier et al. 2011). The RAS proteins are small GTPases that transduce signals within the MAPK/ERK pathway, essential for cell proliferation and survival. Oncogenic mutations in codons 12, 13 or 61 stabilize the Ras-GTP complex, resulting in a gain of function (reviewed in Prior et al. 2012). Despite the high degree of homology between the three genes, *KRAS* mutations are more common in human cancers in general. In UBC, *HRAS* is mutated in 5.5% of tumors, *KRAS* in 5.5%, and *NRAS* mutations are rare (Fernández-Medarde and Santos 2011; Kompier et al. 2011). Interestingly, HRAS has been shown to have a stronger effect on PI3K pathway activation than KRAS and NRAS (Yan et al. 1998; Li et al. 2004), suggesting that *HRAS* and *PIK3CA* mutations might fulfill the same function in the non-aggressive pathway leading to bladder tumor formation.

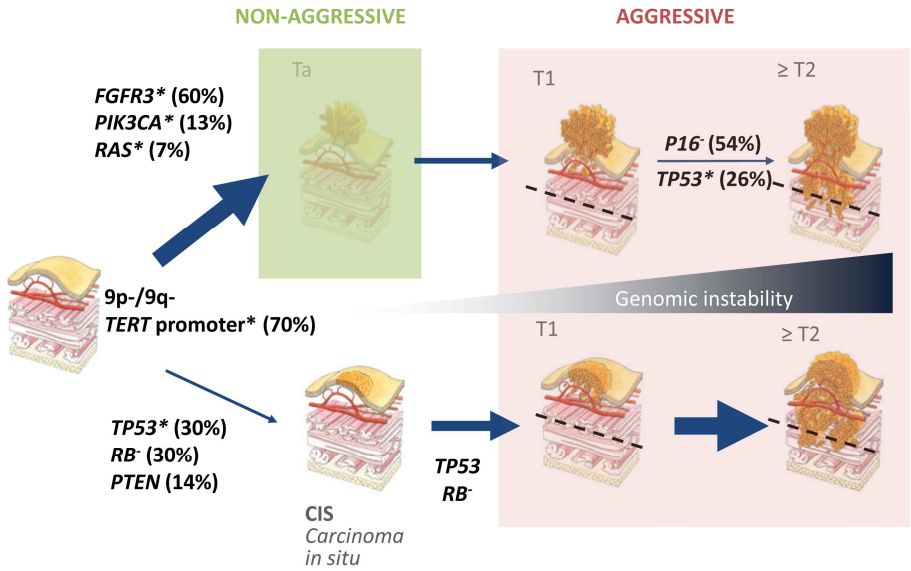

Figure 3. Molecular pathways of UBC development and progression. The non-aggressive pathway (upper) exhibits mutations in oncogenes and is genomically stable, whereas the aggressive pathway (lower) displays alterations in tumor supressors and is unstable. *TERT* promoter mutations and chromosome 9 losses are common to both pathways.

RAS and *FGFR3* mutations are mutually exclusive (Jebar et al. 2005; Platt et al. 2009; Kompier et al. 2011), in agreement with the fact that both proteins effect on ERK1/2 signaling (di Martino et al. 2009). Overall, mutations in *RAS* genes do not associate with grade or stage (Kompier et al. 2011) but a study reported a robust association between low *HRAS* mRNA expression and higher risk of progression in Ta UBC (Birkhahn et al. 2010). Further studies are needed in order to clarify the prognostic potential of mutations and expression levels of *RAS* genes.

6.3. The G1-S cell cycle checkpoint and progression to aggressive tumors

The G1-S checkpoint of the cell cycle is of particular importance for bladder cancer, through frequent alterations of *CDKN2A*, *TP53*, and *RB*. Cell cycle progression is controlled by cyclin-dependent kinases (CDKs) and cyclins, which interact with CDKs, regulating their activity. Cyclin-CDK complexes are formed and dissociated as the cell cycle advances, phosphorylating target proteins that permit cell cycle progression (reviewed in Malumbres and Barbacid 2009). Cyclin D1-3 bind to CDK4 and CDK6 for entry into G1, a stage of the cycle which also requires cyclin E binding CDK2 in order to progress into S phase (Malumbres and Barbacid 2009).

In G1, the retinoblastoma protein (Rb) is bound to the transcription factor E2F; this binding represses E2F activity which is essential for the G1-S transition to take place. Cyclin D-CDK4/6 complexes phosphorylate Rb, releasing E2F and allowing cell cycle progression into S phase. CDK activity is also regulated by the CDK inhibitors (CDKis). The CDKi p16 (encoded by *CDKN2A*) blocks cyclin D-CDK4/6 binding; this prevents Rb phosphorylation and inhibiting progression into S phase. P53 also acts as a tumor suppressor at the G1-S checkpoint, inducing G1 arrest. In normal cells, p53 is expressed at low levels. Upon sensing DNA damage or other types of stress, p53 is stabilized and activated leading to p21 (encoded by *CDKN1A*) upregulation, which prevents the formation of cyclin-CDK complexes by binding to cyclins (Malumbres 2001; Malumbres and Barbacid 2009). Additionally, p53 represses c-Myc expression in a p21-independent manner (Ho et al. 2005). Both p21 upregulation and c-Myc repression are required for p53-dependent arrest of cells in G1, allowing DNA repair to occur. Therefore, Rb, p16, p21, and p53 are all negative regulators of G1-S transition, and thus loss of their normal function results in deregulated cell cycle progression. Alterations in all these tumor suppressors have been linked to UBC progression.

6.3.1. *CDKN2A/*p16

CDKN2A mRNA is undetectable in normal urothelium but is overexpressed in a subset of UBC, with levels increasing in tumors of higher stages and grades (Le Frere-Belda et al. 2004). Stage T1

tumors exhibit loss of p16 protein expression in 54% of the cases, and this loss is associated with an increased likelihood of tumor progression (Krüger et al. 2005). Losses of 9p21, where *CDKN2A* maps, are common in UBC (Williamson et al. 1995). Moreover, a recent report showed that *CDKN2A* homozygous deletions are more common in *FGFR3*-mutant tumors, regardless of their degree of muscle invasion (Rebouissou et al. 2012). Additionally, NMI tumors mutant for *FGFR3* that have lost one or both copies of *CDKN2A* were associated with an increased risk of progression (Rebouissou et al. 2012). These findings strengthen the hypothesis that *CDKN2A* loss is involved in the progression of NMI bladder tumors arising through the non-aggressive pathway.

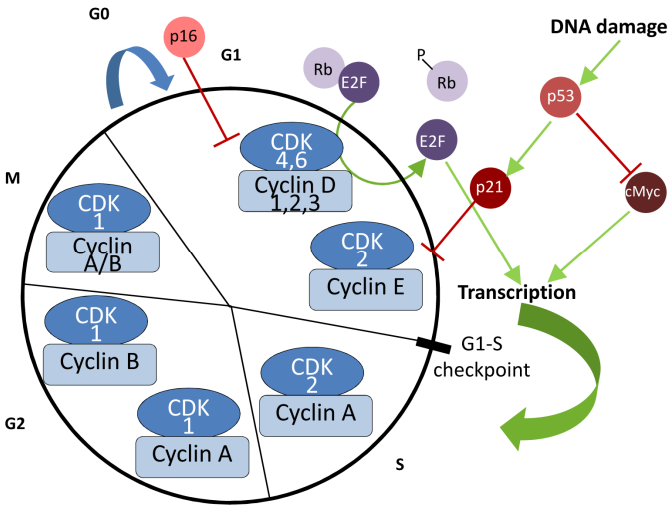

Figure 4. The G1-S checkpoint of the cell cycle. P16 represses cyclin D-CDK4/6 complex formation; this prevents Rb phosphorylation, so E2F is not released and cannot transcribe its target genes. Additionally, DNA damage induces p53 expression, which represses c-Myc dependent transcription and upregulates p21; p21 impedes cyclin-CDK complex formation by binding to cyclins. These concerted processes arrest cells in G1. Upon alterations in any of the components, the G1-S checkpoint is bypassed.

6.3.2. *TP53*

Missense mutations in the *TP53* tumor suppressor gene are the most frequent alterations in human cancers (Olivier et al. 2010). Mutations in *TP53* have also been suggested to lead NMIUBC to progress and invade muscle. *TP53* mutations occur more frequently in the highly conserved DNA-binding domain (codons 125-300), abrogating its transcriptional activity (Olivier et al. 2010). The majority of alterations in this domain are point mutations, most of which confer the altered protein a longer half-life, causing its nuclear accumulation (reviewed in Lane 2004). Therefore, immunohistochemical staining is frequently used as a surrogate for *TP53* mutations. More than 1/4 of NMI tumors exhibit abnormal nuclear staining of p53; these patients are at a higher risk of recurrence than those with normal staining (Chatterjee et al. 2004a). Overexpression of p53 in T1 tumors is associated with a higher risk of progression (Sarkis et al. 1992; Grossman et al. 1998). Concurrent alteration of p16 and p53 expression in Ta and T1 tumors results in a significantly increased stage and grade-adjusted risk of progression of NMIUBC (Hitchings et al. 2004).

6.4. The aggressive pathway

Although CIS are flat lesions, they exhibit aggressive features and often progress to MI bladder tumors (Cheng et al. 1999). CIS present frequent mutations in the tumor suppressor genes that control G1-S progression, making these alterations characteristic of the aggressive proposed mechanism of bladder tumor development. The genetic instability resulting from those mutations results in the accumulation of numerous genetic abnormalities in MIUBC.

6.4.1. *TP53*

Overall, mutations in *TP53* have been detected in about 30% of high-grade UBC (Spruck et al. 1994; Malats et al. 2005). In mouse models, loss of p53 in combination with either *Rb* or *Pten* deletions has been shown to induce CIS and favor progression into MI tumors (Zhang et al. 1999; Puzio-Kuter et al. 2009). In agreement with this observation, CIS exhibit common *TP53* alterations (Hartmann et al. 2002). Independently of the stage and grade of lesions, mutations in *TP53* have been related to a poor prognosis (Esrig et al. 1994). However, a systematic meta-analysis highlights the fact that limitations associated with testing for p53 overexpression is not a robust independent marker to predict outcome in the clinical practice (Malats et al. 2005).

6.4.2. *RB*

Genetic alterations in the *RB1* locus occur in approximately 30% of MI tumors (Miyamoto et al. 1995), and *RB1* LOH has been shown to mainly associated with MI tumors (Cairns et al. 1991). Both mutations and LOH in *RB1* result in loss of protein expression. However, overexpression of Rb has also been detected in UBC (Cote et al. 1998) due to the increased expression of a hyperphosphorylated form of pRB, thought to be caused by p16 loss and/or cyclin D1 overexpression (Chaterjee et al. 2004b). This hyperphosphorylation inactivates the RB tumor suppressor pathway. Therefore, inactivation of the RB pathway in MIUBC occurs through LOH or mutations of the *RB* gene, or by RB protein hyperphosphorylation. As mentioned above, in preclinical models this inactivation cooperates with p53 in the development of MIUBC.

6.4.3. *PTEN*

LOH or homozygous mutations leading to *PTEN* loss of function or decreased protein expression have been detected in 14% of aggressive UBCs (Cairns et al. 1998). PTEN is a lipid phosphatase that negatively regulates the PI3K pathway through dephosphorylation of PIP3, producing PIP2 (reviewed in Knowles 2006). Therefore, decreased PTEN expression leads to

higher proliferation and lower apoptosis by overactivation of PI3K signaling. Alterations of *PTEN* in UBC include point mutations and, more frequently, LOH. An inverse relationship between *FGFR3* mutations and *PTEN* loss has been observed (Platt et al. 2009). This pattern can be explained by the fact that both *PTEN* loss and *PIK3CA* activating mutations lead to PI3K pathway activation. The mechanisms by which *PTEN* and *PIK3CA* alterations happen specifically in one of the pathways are not understood, although they could be related to the higher frequency of genomic instability in

Table 2. Summary of genetic changes found in UBC. The main genes implicated in cancer described in each region appear in brackets. Extensively modified from Knowles 2008.

Deletions			Amplifications			Gains		
Cytogenetic location	Tumor stage	Frequency (%)	Cytogenetic location	Tumor stage	Frequency (%)	Cytogenetic location	Tumor stage	Frequency (%)
2q	Ta, ≥T2	10, 12	1q22	≥T2	<5	1q	Ta, ≥T2	11-14, 17-33
5q	≥T2	15-24	3p24	≥T2	<5	3q (*PIK3CA*)	≥T2	18
6q	≥T2	15-28	6p22 (*E2F3*)	≥T2	5 - 10	5p (*TRIO*)	≥T2	24-37
8p (*FRP1*)	Ta, ≥T2	16, 29-34	8p12	Ta, ≥T2	occasional, <5	7p (*EGFR1*)	≥T2	20
9p (*BNC2, CDKN2A/B, PTPRD*)	Ta, ≥T2	36-47, 21-30	8q21-22 and q24 (*MYC*)	≥T2	<5	8q (*MYC*)	≥T2	23-34
9q (*DAPK1, DBC1, PTCH, TSC1*)	Ta, ≥T2	44-66, 17	10p13-14	≥T2	<5	10p	≥T2	12
10p	Ta	20	11q13 (*CCND1*)	Ta	occasional	17q (*ERBB2, TOP2A*)	Ta, ≥T2	14, 30
10q (*PTEN*)	Ta, ≥T2	20, 16-21	12q15	≥T2	<5	20p	≥T2	21
11p (*KRAS; KAI*)	Ta, ≥T2	10-24, 18-24	17q21 (*ERBB2, TOP2A*)	≥T2	<5	20q (*AURKA*)	Ta, ≥T2	13-17, 26-28
11q	≥T2	22	20q13 (*AURKA*)	≥T2	<5			
13q (*RB1, P33ING1*)	Ta, ≥T2	17, 19						
15q (*RAD51*)	≥T2	13						
16q (*E-CAD*)	≥T2	15						
17p (*TP53*)	Ta, ≥T2	15, 17-24						
18q	Ta, ≥T2	13, 16-17						
Y	Ta, ≥T2	24-28, 21						

the aggressive pathway. In fact, LOH of 10q23, where *PTEN* maps, is more common in aggressive than in non-aggressive bladder tumors (Aveyard et al. 1999; Wang et al. 2000; Harris et al. 2008). Moreover, the expression signature associated to loss of PTEN is a significant predictor of poor outcome in UBC (Saal et al. 2007). The collaboration between Pten and p53 in MIUBC formation

observed in mice (Puzio-Kuter et al. 2009) could be explained by wild type Pten's role in inhibiting invasion in UBC (Gildea et al. 2004).

6.4.4. MIUBC and genomic instability

There is extensive evidence indicating that non-aggressive UBCs are genomically stable whereas aggressive UBCs are unstable. Array CGH showed that the higher the grade and stage of tumors, the greater the number of chromosome alterations (Blaveri et al. 2005). More than half of MI tumors present concurrent alterations in *TP53* and *RB* (Cote et al. 1998; Grossman et al. 1998). Tumors with defects in both genes present worse outcome than tumors harboring mutations in one of them only. Moreover, alterations in either or both of these genes were recently found to correlate with increased copy number changes in high-grade bladder tumors (Iyer et al. 2013). *TP53* mutations have been proposed to contribute to the escape of apoptosis in aneuploid tumors (Li et al. 2010). MI tumors with alterations in *TP53* exhibit greater chromosome instability (Lindgren et al. 2010). Combined analysis of gene expression and genomic profiling has shown that Ta tumors constitute a different group than MI and NMI T1 tumors, supporting the classification of UBC into non-aggressive and aggressive cases (Lindgren et al. 2010). A summary of the most common genetic alterations in UBC detected by CGH can be found in Table 2.

7. UBC AND EPIGENETIC ALTERATIONS

Aberrant methylation is common in many tumor types, including UBC. Promoter hypermethylation at CpG islands has been shown to silence gene expression in cancer (reviewed in Baylin and Jones 2011). Aberrant methylation is observed both in loci presenting genetic alterations, such as *CDKN2A*, and in otherwise unaffected genes like *DAPK* (Chan et al. 2002). These abnormal methylation patterns have also been shown to correlate with grade, stage, and prognosis (Maruyama et al. 2001). A recent study described the existence of 7 chromosomal regions concurrently silenced by methylation of H3K9 and H3K27 and/or H3K9 hypoacetylation, mainly in CIS and MI tumors (Vallot et al. 2011), suggesting a common mechanism of epigenetic regulation affecting multiple genomic regions in a coordinated manner. There is some debate as to whether this mechanism involves only histone modifications or also implies cytosine methylation (Frigola et al. 2006; Vallot et al. 2011). The histone modifiers or the chromatin remodelers responsible for the epigenetic alterations in UBC have not been identified yet. Of note, altered DNA methylation in leukocyte DNA has also been found in patients with UBC, the relationship with tumor methylation being unknown (Moore et al. 2008).

8. EXOME SEQUENCING OF UBC

Even though our knowledge about the events underlying UBC development, and progression has improved greatly over the last decade, the genetic alterations responsible for this tumor are incompletely defined. In the last few years, the advent of next generation sequencing (NGS) has led to a quantum leap in the knowledge of the genetic landscape of human cancer (Alexandrov et al. 2013; CGARN et al. 2013; Puente and López-Otín 2013). Whole exome sequencing (WES) is a less expensive, informative, alternative to whole genome sequencing studies for the purpose of gene discovery in complex diseases such as cancer (Stratton et al. 2009). The first 9 UBC exomes were sequenced in 2011 (Gui et al. 2011). This, together with a series of WES and whole genome sequencing (WGS) studies published within the last year, have helped us identify new pathways involved in UBC (Guo et al. 2013; Cazier et al. 2014; CGARN 2014a; Nordentoft et al. 2014). I have focused my work on two genes involved in the chromatin remodeling pathway (*ARID1A*) and the chromosome segregation pathway (*STAG2*). The findings extracted from the NGS studies, together with our work and other targeted sequencing studies will be explored in the discussion (Solomon et al. 2013; Kim et al. 2014; Ross et al. 2014; Taylor et al. 2014).

9. *ARID1A*

9.1. The BAF SWI/SNF chromatin remodeling complex

SWI/SNF complexes remodel chromatin in an ATP-dependent way. They are thought to mobilize nucleosomes in a multi step process that starts with their binding to DNA at the nucleosomes, which hinders the contact between histones and DNA (Wilson and Roberts 2011). This is followed by ATP-dependent translocation of DNA through the formation of DNA loops. The sliding is propagated along the nucleosome to increase the accessibility of chromatin-repressed DNA binding sites. In mammals, there are two multiprotein complexes, BAF and PBAF, which differ in their subunit composition. BAF complexes contain three highly conserved catalytic subunits: BAF155, BAF170, and SNF5. BAF complexes include an ATPase subunit that can be either BRG1 or BRM (Wilson and Roberts 2011). There are six additional variant subunits, most of which have one or more paralogs that are thought to contribute to target specificity.

9.2. The ARID1A subunit of BAF

One of the non-catalytic variant subunits of BAF is either the AT-rich interactive domain-containing protein 1A (ARID1A) or its mutually exclusive paralog ARID1B. The ARID proteins have a DNA-binding domain (ARID domain) that, in the case of ARID1A, binds DNA non-specifically

(Wilsker et al. 2004). The C terminus of ARID1A is believed to engage in protein-protein interactions. The specific role of ARID1A within the BAF complex has not been elucidated, but it is thought to contribute to specification of target recruiting (Wu and Roberts 2013).

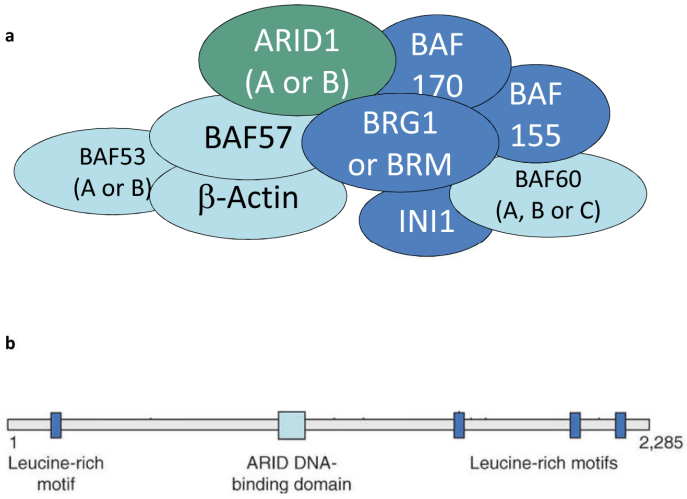

Figure 5. The BAF SWI/SNF Complex and ARID1A. (a) Diagram portraying core subunits in dark blue, ARID1 in green, and other members of the complex in light blue. (b) ARID1A protein scheme, showing its known functional domains. Adapted from Le Gallo et al. 2012.

10. STAG2

10.1. The cohesin complex

Cohesin is a ring-shaped multiprotein complex comprised of four subunits. In somatic cells, these subunits are SMC1A, SMC3, RAD21, and either STAG1 or STAG2 (Remeseiro and Losada 2013). SMC1A and SMC3 are ATPases that modulate chromosomal structure. RAD21 can be cleaved, causing cohesin disassembly, whereas STAG1 or STAG2 are thought to confer functional specificity. When the ring needs to dissociate from chromatin, PDS5A or B and WAPL associate to the complex, allowing its unloading from DNA.

Cohesin binding to DNA is cell-cycle regulated (Losada 2014). The ring is loaded onto DNA by the NIBPL-MAU2 heterodimer at G1. Initially, it only encircles one chromatid but, upon DNA duplication during the S phase, the complex surrounds both sister chromatids and becomes cohesive through acetylation of SMC3 by ESCO1/2 and remains bound to DNA during G2. During prophase, STAG2 is phosphorylated by PLK1, leading to disassembly of the complex. This does not happen at the centromeres, where the SGO1-PP2A dimer protects STAG2 from phosphorylation, allowing chromosome alignment at the metaphase plate. Centromeric cohesin remains assembled until anaphase, when RAD21 is cleaved by ESPL1 (separase), unloading all remaining cohesin and allowing for chromosome segregation. A failure in any of these steps would lead to incorrect chromosome segregation and hence to aneuploidy in the daughter cells.

10.2. Functions of the cohesin complex

The canonical function of the cohesin complex is to regulate sister chromatid cohesion. This is essential not only to guarantee proper chromosome segregation but also to ensure appropriate DNA repair mediated by homologous recombination in response to DNA damage (Remeseiro and Losada 2013). Recent studies have unraveled additional roles of cohesin through DNA-looping, in collaboration with additional factors such as the insulator CTCF (Losada 2014). This affects the three-dimensional structure of DNA, playing a role in transcriptional regulation of gene expression, facilitation of genetic recombination in the immune system, and organization of DNA replication factories widening the functions of cohesin.

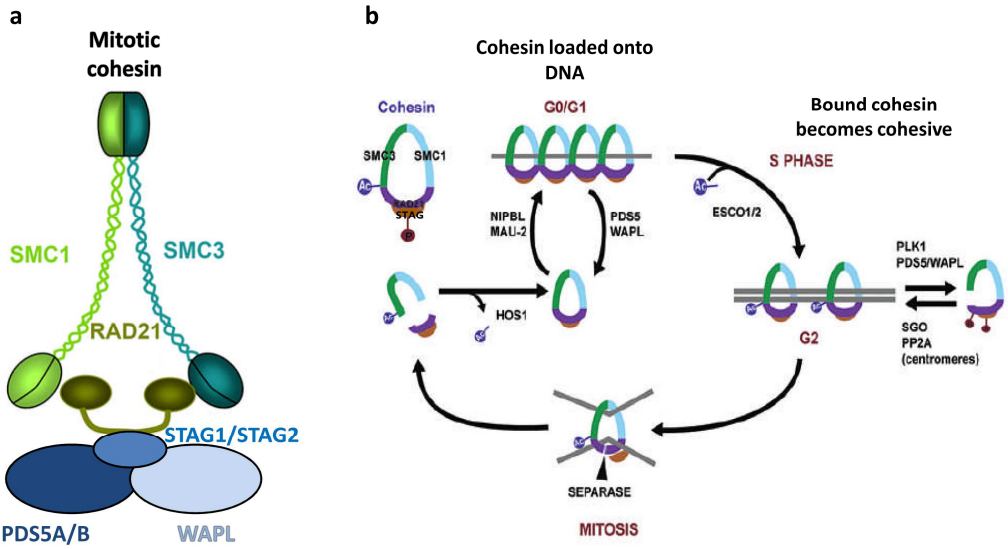

Figure 6. Binding of mitotic cohesin to DNA is cell cycle regulated. (a) Composition of the multiprotein mitotic cohesin complex. Modified from Kitagawa 2009. (b) Cell cycle regulation of cohesin binding to chromatin. Adapted from Rhodes et al. 2011.

10.3. STAG2 function within mitotic cohesin

The distinct function of STAG2 within the cohesin complex is not well understood. Studies in mouse models and human cell lines have shown that STAG1-containing rings are responsible for cohesion at telomeres, whereas STAG2-containing complexes specifically promote cohesion at telomeres (Canudas and Smith 2009; Remeseiro 2012a). Regarding transcriptional regulation, STAG1 has been shown to influence gene transcription in mouse embryonic fibroblasts (MEFs) and mouse brain, a role that STAG2 cannot fulfill (Remeseiro 2012b). Additional studies are required in order to determine the specific functions of each of the cohesin components within the complex, as well as to elucidate how such specificity is attained.

Objectives

Bladder cancer encompasses a heterogeneous group of tumors with diverse clinical, histopathological, biological, and molecular features. Clinical and genetic evidences support the notion that there are at least two genetic pathways leading to urothelial bladder cancer but our knowledge of the genes involved in those pathways remains incomplete. Improving our understanding of the molecular players involved in this disease would contribute to a better classification, assessment of prognosis, and selection of treatment of bladder cancer patients. Therefore, the main goal of this thesis was to identify novel genes involved in bladder cancer development and progression, place them in the context of the current knowledge, and investigate their role in urothelial biology, and studying their potential clinical relevance.

The specific objectives of this PhD thesis were:

1. To identify novel genes mutated in bladder cancer, with a focus on NMIUBC.

2. To describe the frequency and type of mutations found in recurrently altered genes (*ARID1A* and *STAG2*) and assess their relationship with tumor stage and grade, the known pathways of bladder tumor development, and patient outcome.

3. To determine the mechanisms of *STAG2* inactivation in bladder cancer and the contribution of said inactivation to tumor development and progression.

Objetivos

El cáncer de vejiga abarca un grupo heterogéneo de tumores, con características clínicas, histopatológicas, biológicas y moleculares variadas. Las evidencias clínicas y genéticas apoyan la noción de que existen al menos dos vías genéticas que conducen al cáncer urotelial, pero nuestro conocimiento sobre los genes involucrados en dichas vías sigue siendo incompleto. La mejora de nuestra comprensión de los actores moleculares involucrados en esta enfermedad contribuiría a mejorar la clasificación, valoración del pronóstico y selección del tratamiento de los pacientes con cáncer de vejiga. Por tanto, el principal objetivo de esta tesis fue identificar nuevos genes involucrados en el desarrollo y la progresión del cáncer de vejiga, situarlos en el contexto de lo que ya se conoce, y describir su función en la biología del urotelio.

Los objetivos específicos de esta tesis doctoral fueron:

1. Identificar nuevos genes mutados en cáncer de vejiga, centrándonos en los tumores no invasivos.
2. Describir la frecuencia y tipo de mutaciones encontradas en genes alterados de forma recurrente (*ARID1A* y *STAG2*) y valorar su relación con estadio y grado del tumor, las vías conocidas de desarrollo de tumores de vejiga y los resultados de los pacientes.
3. Determinar los mecanismos de inactivación de *STAG2* en cáncer de vejiga y la contribución de dicha inactivación al desarrollo y progresión tumoral.

Materials and methods

1. PATIENTS

Cases from two different series were included for exome sequencing and prevalent mutation screenings. The first cohort belongs to the Integrated Study of Bladder Cancer (ISBLAC), a study conducted between 2008-2012 at Hospital del Mar, Barcelona. The second is part of the EPICURO/Spanish Bladder Cancer Study (SBCS) (Hernández et al. 2006). Cases from the latter were also used for protein expression analysis by immunohistochemistry (see below). The main characteristics of the patients from these cohorts included in this work are summarized in Supplementary Tables 1-3.

Grading, staging, and follow-up of the patients were performed as previously described (Hernández et al. 2006). Diagnostic slides from all tumors used for STAG2 protein expression detection by immunohistochemistry were reviewed by expert pathologists. The NMIUBC group comprised TaG1 and TaG2 tumors (low risk) and TaG3, T1G2, and T1G3 tumors (high risk). The MIUBC group included all tumors ≥T2. A more detailed description of subject characteristics can be found in Supplementary Table 2. For certain analyses, tumors were classified as "non-aggressive" (comprising the low risk NMIUBC) and "aggressive" (comprising the high risk NMIUBC and the MIUBC). Tumors were grouped in this manner because there is extensive evidence indicating that high risk NMIUBC share the molecular/genetic characteristics of MIUBC and likely represent MIUBC tumors that have not yet invaded muscle.

Subjects with NMIUBC were assessed for recurrence and progression. Recurrence was defined as the reappearance of a NMI tumor after a negative monitoring medical evaluation. NMIUBC progression was defined as the development of a MI tumor from a NMI one. Subjects with MIUBC were assessed for progression and death. In this case, progression was defined as the development of new local or metastatic tumors after the patient had received primary treatment. For subjects with MI tumors, all deaths were registered but only those related to UBC were taken into account in survival analyses. Survival periods comprise the time elapsed between diagnosis and the either the last follow-up or the time of death. Written informed consent was provided by all patients. All studies were approved by the ethics committees of all institutions involved.

2. CELL LINES

Cell lines used in this thesis had the following origins: JON, MGH-U4, RT4, SCaBER, SW-800, SW-1710, T24T, VM-CUB-2, VM-CUB-3, 253J, 639V, and CM5637 were purchased from the American Type Culture Collection (Rockville, MD, USA); UM-UC-2 was provided by D. Theodorescu

(University of Colorado, Aurora, CO, USA); UM-UC-3, UM-UC-4, UM-UC-5, UM-UC-6, UM-UC-7, UM-UC-9, UM-UC-10, UM-UC-11, UM-UC-12, UM-UC-13, and UM-UC-18 were provided by H. B. Grossman (MD Anderson Cancer Center, Houston, TX, USA) (Sabichi 2006); J82, MGH-U3, and RT112 cells were provided by F. Radvanyi (Institut Curie, Paris, France); 92-1, 96-1, 97-1, and 97-7 were generated by C. Reznikoff (Yeager et al. 1998) and provided by M. Knowles (University of Leeds, Leeds, UK); and LGWO1 G600 was provided by J. Reeder (U. Rochester, NY, USA). All cells were periodically tested to make sure they were not contaminated with Mycoplasma.

Cells were grown in Dulbecco's Modified Eagle's Medium (DMEM, Sigma) supplemented with 10% heat-inactivated fetal bovine serum (FBS, Sigma) and 1 mM sodium pyruvate (Invitrogen).

3. GENOMIC DNA EXTRACTION

Sections from formalin-fixed paraffin-embedded tumor blocks were microdissected and genomic DNA was extracted as described elsewhere (Hernández et al. 2005).

4. RNA ISOLATION

For reverse transcriptase and quantitative PCR (RT-qPCR), cells in log-growth phase were washed with Phosphate Buffer Saline (PBS) (Sigma) and total RNA was isolated with the GenElute Mammalian Total RNA kit (Sigma). For RNA-sequencing, RNA from control and STAG2 knock-down UBC cells was extracted and purified using the phenol-chloroform method. RNA integrity was assayed on an Agilent 2100 Bioanalyzer (range 6.8-9.1).

5. EXOME SEQUENCING

The Agilent SureSelect Human All Exon plus capture kit was used for the preparation and enrichment of libraries. V3 50 Mb was used for samples 114, 116, 193, 251, 310, 331, 413, 418 and Esp66, whereas the libraries for samples 062, 064, 179, 188, 274, 313, 343 and 451 were prepared using v4 51Mb. The libraries were hybridized to an Illumina flow cell and samples were sequenced on a HiSeq 2000 instrument, obtaining 75-bp paired-end reads.

The Illumina Real-Time Analysis pipeline was used for base calling and quality control. The Genome Multitool Mapper was initially used to trim sequence reads. The limit was set at the first base with >10 mapping to the human genome build hg19 (GRCh37). The 27 (~4%) reads excluded from this process were subjected to a second round of mapping using BFAST. Non-duplicated, uniquely mapping, read pairs were selected and results from both mapping rounds were merged. Single nucleotide variants (SNVs) and short indels were called using the default settings of the

SAMtools suite version 0.1.18. SNVs falling within regions exhibiting read depth <10, having low mappability, strand bias P < 0.001, or tail distance bias P < 0.05 were excluded. The results from blood and tumor exome sequencing were compared to call for somatic mutations. Variant-supporting read counts were used to perform Fisher's exact tests. Only those somatic SNVs with Fisher's exact test P value <0.0001 were considered.

5. 1. SNV damage prediction

SNVs leading to non-synonymous changes or mapping to exon junctions were classified as 'relevant'. Variants causing amino acid substitutions were assessed for damaging potential using the scores of the SIFT (Kumar et al. 2009) and MutationAssessor (Reva et al. 2011) publicly available tools, which were normalized to fall within a 0-1 range. MutationAssessor 'low risk of damage' predictions were assigned a value of 0.5, 'medium risk of damage' was given 0.7, and 'high risk of damage' was assigned a value of 1. Whenever both predictions could be obtained, an average score was calculated; if one of the tools did not yield a prediction, the score from the other one was used. SNVs were considered damaging when the final score was >0.8. Relevant exon junction SNVs were classified as damaging if they fell within the first 2 bases of the exon or 8 bases of the intron, counting from donor junctions, or if they fell within 3 bases into the exon or 8 into the intron, counting from acceptor junctions. Frameshifts and stop gains ablating >30% of a protein sequence or whole protein domains (as annotated in InterPro) were labeled as damaging. P values for the genes associated to the SNVs were calculated combining scores from both methods, combining recurrence and functional impact data as proposed in the Oncodrive-fm approach (González-Pérez and López-Bigas 2012).

Mann-Whitney U test was used for statistical analyses comparing the ratio of non-synonymous vs. synonymous mutations across different age groups, tumor stages, and smoking status. The frequency of mutations of recurrently altered genes in non-aggressive vs. aggressive tumors was compared using Fisher's exact tests.

5.2. Pathway analysis

Pathway analysis was performed as outlined elsewhere (Vázquez et al. 2012). Three different gene lists were compiled: 1) Relevant genes (those with relevant mutations, see above); 2) Damaged genes (those with damaging mutations, see above); 3) Recurrently altered genes (those presenting relevant mutations in at least 2 samples). These lists were analyzed for enrichment based on a hypergeometric test, taking into account KEGG pathways and GO (Gene Ontology

biological process) components, correcting for genes present in different clusters. Benjamini-Hochberg's False Discovery Rate (FDR) scores were used to adjust P values and only FDR < 0.1 were taken into account.

6. PREVALENCE SCREENING OF RECURRENTLY MUTATED GENES

6.1. HaloPlex targeted resequencing

Targeted resequencing was performed for five of the cases included in the exome sequencing discovery screen, to validate and compare both experimental strategies, and for 60 additional cases. The HaloPlex Target Enrichment System (Agilent) was used to capture the exonic regions of the 110 genes of interest (see Supp. Table 4), following manufacturer's instructions. Briefly, restriction digestion was performed to create a library of DNA fragments. This library was then hybridized with probes targeting regions of interest that incorporated Illumina paired-end motifs and index sequences. This was followed by DNA capture and PCR amplification using Herculase II Fusion Enzyme or KAPA HiFi HotStart polymerase. AMPure XP beads were used to amplify PCR products. Illumina multiplex sequencing was performed on the purified amplicons. The Flexible Adapter Remover program was used to eliminate adaptor and primer sequences from the sequencing reads.

Reads were mapped after eliminating adaptor and primer sequences. Indels and SNVs were called as described above; duplicates were included, but variants with strand bias P < 0.001 or tail distance bias P < 0.05 were excluded. SNVs falling outside the selected sequences, displaying low mappability or read depth <10, annotated as SNPs in the 1000 Genomes Project (21 May 2011 version), or those occurring in >1% of the reads in paired normal blood samples were filtered out. Variants present in tumors were compared with their corresponding blood samples to select somatic mutations; those with Fisher's exact test P value <0.0001 were included. The predicted effect of the mutations was annotated as detailed above.

Frequency of mutations of recurrently altered genes in non-aggressive vs. aggressive tumors was compared using Fisher's exact tests. All analyses were performed using the R 2.15.1 statistical software. P values <0.05 were required for observations to be considered statistically significant.

6.2. ARID1A mutational analysis

Concurrently, *ARID1A* mutational analyses were performed using tumor-blood DNA pairs from 52 additional cases from the ISBLAC series and 5 UBC cell lines (MGH-U3, RT112, UM-UC-3, UM-

UC-17, and VMCUB-3), following a protocol described elsewhere (Wiegand et al. 2010). In short, exons 2-20 were amplified using AccuPrime Taq DNA polymerase HiFi (Invitrogen). PCR products from all samples were pooled at equimolar concentrations and the Covaris S2 shearing instrument was used to fragment the amplimers to 100-300 bp. DNA (40-80 ng) from each sample was subject to end-repair and dA-tailing. Subsequently, fragments were ligated to indexed adapters according to the recommendations for DNA sample preparation from the TruSeq kit (Illumina). PCR (n=10 cycles) was used to amplify adapter-ligated libraries. Amplified libraries were multiplexed and subjected to single read sequencing for 38 cycles (Genome Analyzer IIx with SBS TruSeq v5 reagents, Illumina).

Novoalign V2.07.04 (Novocraft, Selangor, Malaysia) was used to independently align sequence tags against the genomic *ARID1A* sequence in the Human February 2009 (GRCh37/hg19) assembly. Those positions that aligned were subjected to a quality filter and mapped combining custom Perl scripts and SAMtools. The SIFT tool (Kumar et al. 2009) was used to predict the functional consequence of all identified variants.

6.3. FGFR3 mutational analysis

Information on the *FGFR3* mutational status of tumors from the EPICURO study was available from previous published studies of our laboratory, as described in detail elsewhere (Hernández et al. 2005; Hernández et al. 2006). Tumors from cases included in the ISBLAC study were analyzed using the SnapSHOT multiplex assay (Hafner et al. 2006), followed by capillary electrophoresis of the fragments.

6.4. Sanger sequencing

Sanger sequencing was used to verify somatic SNVs called by bioinformatic analysis of the exome sequencing (see above). Additionally, Sanger sequencing was employed to assess *STAG2* mutational status in UBC cell lines. Genomic DNA was used to sequence exons 8-10. Overlapping cDNA amplicons comprising *STAG2* exons 3-7, 11-31, and 33-36 were generated and sequenced.

7. GENE COPY NUMBER ANALYSIS

To assess copy number changes, two different platforms were used:

- A subset of tumors from the EPICURO study, including 12 non-aggressive samples for which matching fixed tissue was available, was analyzed using Human 2.0 BAC arrays (UCSF Cancer Center) (Snijders et al. 2004). In short, each array contains triplicate spots of 2,464 BAC clones

uniformly distributed across the genome, giving a final resolution of 1.4 Mb. Genomic DNA (100 ng for tumor and 50 ng for reference DNA) was amplified and fluorescently labeled by random priming. DNA was mixed with 100 mg Cot-1 DNA (Life technologies, Gaithersburg, MD) and precipitated. DNA was resuspended in 60 ml of hybridization solution (50% formamide, 10% dextran sulphate, 4% SDS, 100 mg yeast tRNA, and 2xSSC), denatured at 72 ºC for 10 min, and repetitive sequences were blocked by incubation at 37 ºC for 1 h. Samples were hybridized to the array for 48 h, washed, and mounted. Fluorescence intensities were measured for tumor and normal samples using a charged coupled device camera (Sensys, Photometric) coupled to a 1x optical system. SPOT and SPROC software (Jain et al. 2002) was used to calculate a test intensity/reference intensity single centered log2 ratio for each BAC clone. Only clones with detectable signal in $\geq$ 2 of the triplicate spots, clones with detectable signal in $\geq$ 20% of samples, and those where the triplicate standard deviation was $\leq$ 0.33% were included in the analysis. Frequency of alterations $\geq$ 40% were considered in the assessment of chromosomal-scale genomic changes.

- An independent set of 11 non-aggressive tumors for which we also had available material for immunohistochemistry was analyzed using Illumina HumanHap 1M BeadChip high resolution SNP arrays, which contain >1 million SNP markers. This array includes 17,202 monomorphic probes located in regions of known or suspected copy number variations (CNVs). Manufacturer recommendations were followed for sample preparation and array hybridization. Briefly, DNA was denatured and amplified for 20 h at 37 ºC. Amplified genomic DNA was enzymatically fragmented, precipitated, and resuspended in Illumina's hybridization buffer. Fragments were applied to BeadChips (1-2 samples per bead) and incubated for 16-24 h at 48 ºC. BeadChips were washed and oligonucleotides on the BeadChips were extended in a one-nucleotide elongation reaction with fluorescently labeled oligonucleotides, using captured DNA as template. Fluorescence intensities were measured using the Illumina iScan System, which records high-resolution images of the light emitted from the fluorophore-labeled single-base extension products. Single nucleotide polymorphism (SNP) calling was performed as published (Rothman et al. 2010; Jacobs et al. 2012). Briefly, bead fluorescence intensities were normalized, and genotypes and R values were extracted using the Illumina Beadstudio software (v.3.1.3.0). After normalization of R values, log R ratios were calculated taking the average R values of blood leukocyte samples from EPICURO control individuals as a reference (Rodríguez-Santiago et al. 2010). Log R ratios provide with a quantitative measure of copy number. Log R ratios = 0 correspond to a sample carrying two copies of the DNA region, log R ratio < 0 indicate deletions,

and ratios > 0 correspond to amplification. The waviCGH software was used to obtain copy number calls (Carro et al. 2010).

8. RNA SEQUENCING

Total RNA was denatured to expose poly(A) tails, bound to oligo(dT) cellulose, and poly(A)+ RNA was eluted, purified, randomly fragmented, converted to double stranded cDNA, and processed through subsequent enzymatic treatments of end-repair, dA-tailing, and ligation to adapters as outlined in Illumina's "TruSeq Stranded mRNA Sample Preparation Part # 15031047 Rev. D", to ensure that only the cDNA strand generated during the first strand synthesis is sequenced. The adapter-ligated library was completed by subjecting samples to 10 cycles of PCR with Illumina PE primers. The resulting purified cDNA library of template molecules was applied to an Illumina flow cell for cluster generation (TruSeq cluster generation kit v5) and sequenced on the Genome Analyzer IIx (GAIIx) with SBS TruSeq v5 reagents following manufacturer's instructions (SingleRead 1x40 bases).

Illumina Real Time Analysis software (RTA1.13) was used for image analysis and per-cycle base calling. Conversion to FASTQ format and sequencing alignment with the ELAND algorithm (v2e) was performed using CASAVA-1.8 (Illumina). Quality check was done using fastqc (v0.9.4, Babraham Bioinformatics). Raw reads were aligned to the reference genome hg19/GRCh37 with tophat (version 2.0.4). Gene expression levels (Fragments Per Kilobase of exon per Million fragments, FPKMs) were quantified with cufflinks (version 2.0.2), as annotated in the Ensembl version GRCh37.65.

FPKM correlations and Principal Component Analysis (PCA) clustering of samples were carried out with the R statistical software's (version 2.14.1) functions cor() and prcomp. Differential gene expression analysis was performed with the cuffdiff function included in cufflinks (version 2.0.2). Venn diagrams were created with the freely available software VENNY (accessible at www.bioinfogp.cnb.csic.es/tools/venny) (Oliveros 2007).

9. GENE EXPRESSION ANALYSIS

For gene expression analyses, RNA from 43 UBC tumors and 60 cell lines had been previously assessed using the Human Genome ST1.0 DNA and the U133A array, respectively (both from Affymetrix), as reported elsewhere (Fu et al. 2012). RNA (500 ng) was amplified, labeled, and hybridized to the arrays. Raw expression data from the Human Genome ST1.0 DNA and the U133A arrays was processed as previously described (Fu et al. 2012; Zieger et al. 2005). Briefly,

the ST1.0 DNA expression values were normalized using the Frozen Robust Multiarray Analysis method (McCall et al. 2010) before conducting further analyses. The U133A expression values were normalized using the GC-RMA procedure implemented in the ArrayAssist software (Stratagene, La Jolla, CA) and they were log2 transformed before analysis. Affymetrix control probes were removed from further assessment.

9.1. Analysis of publicly available gene expression datasets

Normalized data for studies GSE89 (Dyrskjøt et al. 2003) and GSE32894 (Sjödahl et al. 2012) was downloaded from the Gene Expression Omnibus (GEO) database. The data was pre-processed with GEPAS 4.0 (http://www.gepas.org/), to obtain average values for all probes that mapped within a given locus. The average expression of *ARID1A*, *FGFR3*, and *TP53* was calculated for each group and normalized with respect to the first group. Differentially expressed genes were identified as follows: pre-processed data were subjected to an Anova limma analysis with the POMELO online software (http://pomelo2.bioinfo.cnio.es), and differentially expressed genes were identified using an FDR adjusted P-value <0.5 as significance threshold and performing a t-test on all qualifying samples.

10. HISTOLOGICAL ANALYSIS

10.1. Immunohistochemistry (IHC)

Protein expression in tumors was analyzed using tissue microarrays (TMAs) containing 2 cores representative of the corresponding tumor. The cores were extracted from formalin-fixed paraffin-embedded tissue blocks from patients included in the EPICURO/SBCS. Slides were deparaffinized and rehydrated. Antigen retrieval was performed by boiling for 10 min in 10mM sodium citrate buffer (pH6.0). Samples were then incubated in 3% H_2O_2 in methanol for 30 min. Blocking was performed for 1 h with 2% Bovine Serum Albumin (BSA) (Sigma) in PBS. Sections were incubated with the corresponding primary antibodies, diluted in 1% BSA in PBS, overnight at 4 ºC.

The primary antibodies used and the experimental conditions are indicated in Table 3. It is important to note that two different antibodies, yielding 92% concordant results, were used for STAG2 detection. Sections were washed x3 in PBS for 10 min and they were incubated for 1 h with EnVision+HRP labeled secondary anti-rabbit or anti-mouse antibodies (Dako). Signal detection was performed using the DAB+Chromogen system (Dako) at room temperature for 2 min. Nuclear

counterstaining was performed for 2 min using Carazzi's Hematoxylin solution DC (Panreac); slides were dehydrated and mounted.

Table 3. Primary antibodies used in this dissertation. NA: Not applicable.

Target protein	Species	Information	Immunization peptide	WB dilution	IHC/IF dilution
ARID1A	Mouse	3H2, Abnova	NA	NA	2 µg/mL
	Mouse	M02, clone 3H14, Abnova	NA	1:1000	NA
β -CAT	Mouse	β -catenin-1, Dako	NA	NA	ready-to-use
E-CAD	Mouse	NCH-38, Dako	NA	NA	1:50
FGFR3	Mouse	B-9/sc-13121, Santa Cruz	NA	NA	8 µg/mL
Ki67	Mouse	MIB-1, Dako	NA	NA	ready-to-use
KRT5/6	Mouse	D5/16B4, Dako	NA	NA	ready-to-use
KRT14	Mouse	LL002, Novocastra Laboratories	NA	NA	1:25
KRT20	Mouse	Ks20.8, Dako	NA	NA	ready-to-use
Myosin IIa	Rabbit	Cell Signaling	NA	1:1000	NA
P-CAD	Mouse	Clone 56, B.D. Transduction Laboratories	NA	NA	1:75
p53	Mouse	DO-7, Novocastra	NA	NA	1:200
PDS5A	Rabbit	Ana Losada, CNIO, Madrid (Spain)	CKKAVPAERQIDLQ	0.58 mg/mL	NA
PDS5B	Rabbit		VSTVNVRRRSAKRERR	0.65 mg/mL	1 µg/mL
RAD21	Rabbit		GDQDQEERRWNKRTQQML	0.5 mg/mL	0.2 µg/mL
SMC1	Rabbit		CEMAKDFVEDDTTHG	1.3 mg/mL	1 µg/mL
SMC3	Rabbit		CDLTKYPDANPNPNEQ	0.18 mg/mL	0.2 µg/mL
STAG1	Rabbit		CEDDSGFGMPMF	1 mg/mL	2 µg/mL
STAG2	Rabbit		CDPASIMDESVLGVSMF	0.8 mg/mL	0.5 µg/mL
	Mouse	Clone J-12, Santa Cruz Biotechnology, sc-81852	NA	NA	1:200
REC8 N' K675	Rabbit	José Luis Barbero, CIB, Madrid (Spain)	Last 143 amino acids (human)	1:2000	NA
SMC1β K976	Rabbit		Amino acids 1047-1233 (human)	1:1000	NA
STAG3 K403	Rabbit		Amino acids 626–757 (human)	1:2000	NA
STAG3	Rabbit	HPA049106, Sigma Prestige	NA	1:1000	1:200
UPK3B	Rabbit	HPA010506 Sigma Prestige	NA	NA	1:750
Vinculin	Mouse	Sigma	NA	1:10000	NA

ARID1A and STAG2 expression was scored according to staining intensity (0-3) and percentage of positive cells (0-100%). Histoscores were obtained as the product of both variables. Histoscores were used as the input for unsupervised clustering analysis using the heatmap.2 function of the gplots package within the R 2.15.1 statistical environment.

Categorical data reporting was done using number and percentages. Associations between STAG2 or ARID1A loss of expression/mutation and the main characteristics of patients were determined using the χ2 test, ANOVA, t- test, Mann-Whitney (MW) or Kruskal-Wallis (KW) analyses as appropriate. Associations between STAG2/ARID1A expression and other markers were examined using the χ2 test. Odds ratio (OR) and 95% confidence interval (95% CI) were used to assess the association between STAG2 expression and other categorical variables. Survival data was analyzed using Kaplan-Meier plots. The log-rank test was used to assess differences between curves. Multivariable analysis was performed using Cox proportional hazard models. Adjusting factors included are region, gender, age, T stage, grade, tumor size, number of recurrences, number of affected nodes, metastasis, and treatment, depending on the outcome measured.

10.2. Immunofluorescence

Normal human bladder tissue sections were processed as for IHC, with the addition of 0.5% Triton-X 100 to the 2%BSA blocking solution to improve cell permeabilization. Primary antibodies and concentrations used are indicated in Table 3. Sections were washed and incubated in the dark for 1 h with fluorescently labeled secondary antibodies: Cy3 anti-rabbit (1:200) (Jackson ImmunoResearch Laboratories, Inc.) and Alexa 488 anti-mouse (1:200) (Invitrogen). Then, sections were incubated with DAPI (0.5µg/ml in PBS) for 5 min and mounted with ProLong® Gold Antifade Reagent (Life Technologies).

11. PROTEIN ANALYSIS

11.1. Protein lysate preparation

Cells in log-growth phase were lysed in RIPA buffer (10 mM Tris–HCl, pH7.5, 1 mM EDTA, 1% Triton X 100, 0.1% SDS, 0.1% Na-deoxycholate, 100 mM NaCl) supplemented with Phosphatase Inhibitor Cocktail 3 (Sigma), 0.2 mM orthovanadate, and the cOmplete, Mini, EDTA-free Protease Inhibitor cocktail (Roche). Extracts were sonicated in an ultrasonic water bath (30 pulses), and centrifuged for 20 min at 15 000 g. Protein concentration was quantified by spectrophotometry using the BioRad Protein Assay Solution (BioRad).

11.2. SDS-PAGE western blotting

Protein extracts were denatured and subjected to sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). A discontinuous 4%-7.5% (for immunodetection of cohesion components) or 6% (for ARID1A immunodetection) resolving gel was used to separate 50 µg of total protein. Samples were blotted onto nitrocellulose membranes. Membranes were blocked for 1 h with 5% skim milk (SM) in Tris-buffered saline with Tween (TBS-T) and subsequently washed in TBS-T.

Membranes were incubated for 16 h at 4 ºC with primary antibodies as indicated in Table 3, washed, and incubated for an additional 1 h at 25 ºC with HRP-labeled anti-mouse or anti-rabbit secondary antibodies (Amersham Biosciences). Reactions were developed with the Amersham™ ECL™ western blotting detection reagents (GE Healthcare) and protein bands were visualized on Amersham™ Hyperfilm ECL™ (GE Healthcare). Bands of interest were identified by size comparison against the Dual Color Precision Plus Protein™ Standard molecular weight marker (BioRad).

12. RT-qPCR

DNase treatment (DNAfree, Ambion) was performed for 30 min at 37 ºC, and mRNA was reverse-transcribed (Taqman Reverse Transcription Reagents kit, Applied Biosystems). SYBR GreenER (Invitrogen) was used to perform quantitative PCRs in the 7900H Fast Real Time PCR System (Applied Biosystems), using 20ng of cDNA as template. The following primers were used: *ARID1A* CCCCTCAATGACCTCCAGTA (F, forward) and ATCCCTGATGTGCTCACTCC (R, reverse), *STAG2* AGATACCGTGATGCGATAGC (F) and GGCATCACTATACATCTTCATCC (R), *FYN* AAGGACTCACCGTCTTTGG (F) and GTGTCACTCCTGTTCCTCC (R), *TFEB* ACCCTGAGAGGGAGTTGGAT (F) and CAGGGAGGCTGTGACCTG (R), and *CDK6* ACCTACTTCTGAAGTGTTTGAC (F) and TCCTGGAAGTATGGGTGAG (R). Product quality and specificity was controlled by performing melting curve analyses for each reaction. All reactions were performed in triplicate. Reported expression levels are normalized to individual *HPRT* expression values, detected with the following primers: GGCCAGACTTTGTTGGATTTG (F) and TGCGCTCATCTTAGGCTTTGT (R).

13. *IN VITRO* FUNCTIONAL ASSAYS

13.1. Gene silencing

HEK293 cells were transfected using the calcium phosphate/HEPES technique (Kingston et al. 2003). HEK293 cells were co-transfected with 15µg pSPAX, 5µg VSVG, and 15 µg of either lentiviral GFP or Sigma Mission expressing either lentiviral shNT (non-targeting control) or the corresponding shRNAs targeting STAG2 or ARID1A. Infection was allowed to proceed for 16 h and the medium was replaced with fresh complete DMEM. The virus-containing media was collected 24 h later, filtered (Sartorius stedim 0.45 µm minisart), and used to infect the corresponding UBC cells in the presence of 5µg/ml hexadimethrine bromide polybrene (Sigma). Three rounds of infection of 24 h, 72 h, and 24 h, respectively, were performed. Infected cells were selected by trypsinization followed by exposure to 2µg/ml puromycin (Sigma). Cells were collected or seeded for further experiments after 48 h of puromycin selection.

13.2. Gene overexpression

The human *STAG2* cDNA (b isoform; 1,231 residues) was PCR-amplified from Addgene pEGFP-STAG2 plasmid (ref. 31972) and subcloned into the pLVX-puro lentiviral vector. HEK293 transfection, viral supernant production, and subsequent UBC cell infection and selection were performed as described above.

13.3. Colony formation assays

Puromycin-selected cells ($4x10^3$ for ARID1A assays; $8x10^3$ for STAG2 assays) were plated in triplicate in 6-well (ARID1A) or 12-well (STAG2) plates. Cells were allowed to grow under standard conditions for 4 or 7 days, respectively. After washing with PBS, cells were fixed with methanol at -20 ºC for 30 min and incubated with 0.5% crystal violet in 25% methanol for 10 min at room temperature. Cells were washed with distilled H_2O; growth was measured by eluting crystal violet in 10% acetic acid and recording absorbance at 590 nm using a biophotometer (Eppendorf).

13.4. Metaphase analysis

Puromycin-selected cells were incubated with colcemid (0.1 µg/ml) for 6 h to arrest them at metaphase. After hypotonic treatment with 75 mM KCl at 37 ºC for 15, 25, or 30 min (RT112, 639V, and UM-UC-11 respectively), cells were fixed in methanol:acetic acid 3:1, dropped from a height of $\geq$45 cm onto cleaned glass coverslips to obtain chromosome spreads that were stained by G-banding using standard methods. At least 50 metaphases per condition were examined;

chromosome number and centromeric cohesion defects were counted using an Axioplan II Imaging MetaSystem Microsoft and the Ikaros software (Metasystems, GmbH, Altlussheim, Germany). The Wilcoxon rank-sum test was used to compare chromosome number between control and interfered cells.

Work description

Description of the work carried out by the candidate

Cristina Balbás Martínez designed and performed *in vit*ro functional studies, performed immunohistochemical analysis of tumor samples, and contributed to the design and performance of mutation validation analyses. She participated in the statistical analyses of the data, and performed bioinformatic analyses of histoscores and gene expression datasets. She participated in the discussion of the results and contributed to the writing of the manuscripts included in Annex II. Other laboratory members and collaborators designed and prepared HaloPlex libraries for sequencing, processed and analyzed exome sequencing and targeted resequencing data, and performed statistical analyses.

Results

1. DISCOVERY AND PREVALENCE SCREENS

1.1. Exome sequencing

In order to unveil novel genes mutated in UBC, we performed exome sequencing on 17 tumors and their paired normal leukocyte DNA. At least 70% of the cells in the samples included in the screen were tumoral. The study included tumors of different grades and stages but we focused on NMIUBC cases because other sequencing initiatives, such as The Cancer Genome Atlas Research Network (CGARN) project, are centered on MIUBC tumors (see Supp. Table 1 for patient characteristics). Sequencing coverage metrics are shown in Supplementary Table 5. The average coverage was 79±16X for tumors and 82±18X for leukocytes, and the fraction of exons covered by more than 15 reads was 83.7%.

1.2. Mutation number or type is not significantly associated with tumor aggressiveness, patient smoking status, or age.

Bioinformatics analysis identified 2,927 somatic mutations. Individual tumors exhibited a wide range of somatic mutations (4-360). The average of 169±114 SNVs per tumor (Fig. 7a) falls within the range found for other adult solid tumors (Table 4). Among the identified somatic mutations, 1,263 were predicted to be relevant (leading to non-synonymous changes or mapping to intron-exon junctions) and 798 were predicted to be damaging (combined score >0.8 in a 0-1 scale). The most frequent nucleotide changes were C>T transitions and C>G transversions (Fig. 7b).

Figure 7. Distribution of the SNVs identified in the exome sequencing study. (a) Total number of SNVs per tumor. (b) SNVs classified according to type of nucleotide substitution for each case. (c) Predicted effect of all SNVs identified per tumor.

Table 4. **Mean somatic mutations identified in other exome sequencing studies**.

Publication	Patient number	Tissue	Mean somatic mutations per patient	Mean Non-synonymous	Calling method	Hypermutated patients removed
Lee et al. 2012	32	Rhabdoid cancers	5	4	Firehose (MuTect + Genepattern)	
Pugh et al. 2013	221	Neuroblastoma	23	18	MuTect	
Biankin et al. 2012	99	Pancreatic ductal adenocarcinoma	26	20	SAMtools	
Barbieri et al. 2012	111	Prostate cancer	43	32	Firehose (MuTect + Genepattern)	
Le Gallo et al. 2012	12	Endometrial tumors	44	32	MPG algorithm	1
Quesada et al. 2011	105	Chronic lymphocytic leukemia	45	12	SIDRON	
Banerji et al. 2012	103	Breast cancer	48	37	Firehose (Mutect)	
Guichard et al. 2012	24	Hepatocellular carcinoma	NA	41	NA	
Zang et al. 2012	15	Gastric adenocarcinoma	NA	50	GATK	
Brennan et al. 2013	291	Glioblastoma	74	56	Firehose	
Guo et al. 2013	99	Bladder carcinoma	114	93	VarScan+SAMtools	
This study	**17**	**Bladder carcinoma**	**169**	**74**	**SAMtools**	
Nordentoft et al. 2014	30	Bladder carcinoma	195	79	GATK (MuTect)	
Imielinski et al. 2012	159	Lung adenocarcinoma	216 (median)	167 (median)	Firehose (MuTect + Genepattern)	
CGARN 2014a	130	Bladder carcinoma	302	NA	Firehose (MuTect)	
CGARN 2014b	230	Lung adenocarcinoma	306	237	Firehose (MuTect)	
Kumar et al. 2011	20	Prostate cancer	362	215	SAMtools	3
Thompson et al. 2012	33	Breast cancer	NA	284	GATK	
CGARN et al. 2012	224	Colorectal carcinoma	402	310	MuTect	
Seshagiri et al. 2012	70	Colorectal carcinoma	921	350	GATK	2

We used Sanger sequencing to verify a fraction of the mutations called as somatic. We sequenced a total of 226 amplicons from 13 samples (selected according to availability of material) and we verified 219 of them. Of those, 214 (94.7%) were confirmed to be somatic,

supporting the validity of the bioinformatic pipeline used. Supplementary Table 6 shows details of the mutations verified by DNA sequencing.

The non-synonymous:synonymous (NS:S) ratio was <1 in 15/17 (88%) of the samples (Fig.7c, 8a). The NS:S ratio was not significantly different between non-aggressive vs. aggressive tumors (P=0.87) or non-smokers vs. smokers (P=1). We observed a tendency for patients diagnosed at a later age to exhibit more synonymous mutations but the differences were not statistically significant (P=0.11) (Fig. 8b). This could be due to the relatively low number of tumors sequenced.

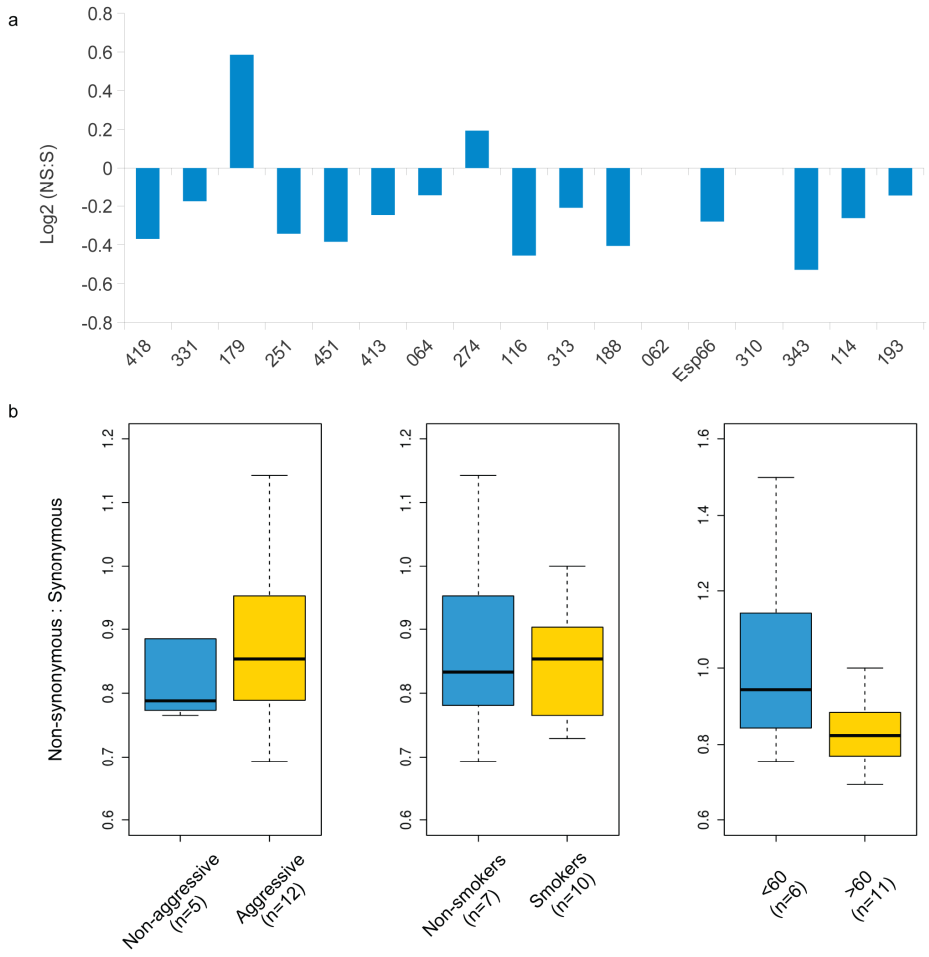

Figure 8. Comparison of non-synonymous (NS) and synonymous (S) variants in the exome sequencing analysis. (a) Log2 values for each tumor NS:S ratios. (b) NS:S comparison according to tumor aggressiveness (P=0.87), smoking habits (P=1), and patient age (P=0.11).

The number of SNVs, indels, transitions, transversions, and mutation subtypes (synonymous; non-synonymous, non-damaging, and damaging) was not significantly different in non-aggressive vs. aggressive tumors (Fig. 9a). The same was true when comparing those variables between non-smoking and smoking patients, although smokers exhibited more damaging mutations than non-

smokers (Fig. 9b). Similarly, patients diagnosed at >60 years old exhibited more alterations, regardless of mutation type, but this difference was not statistically significant (Fig. 9c).

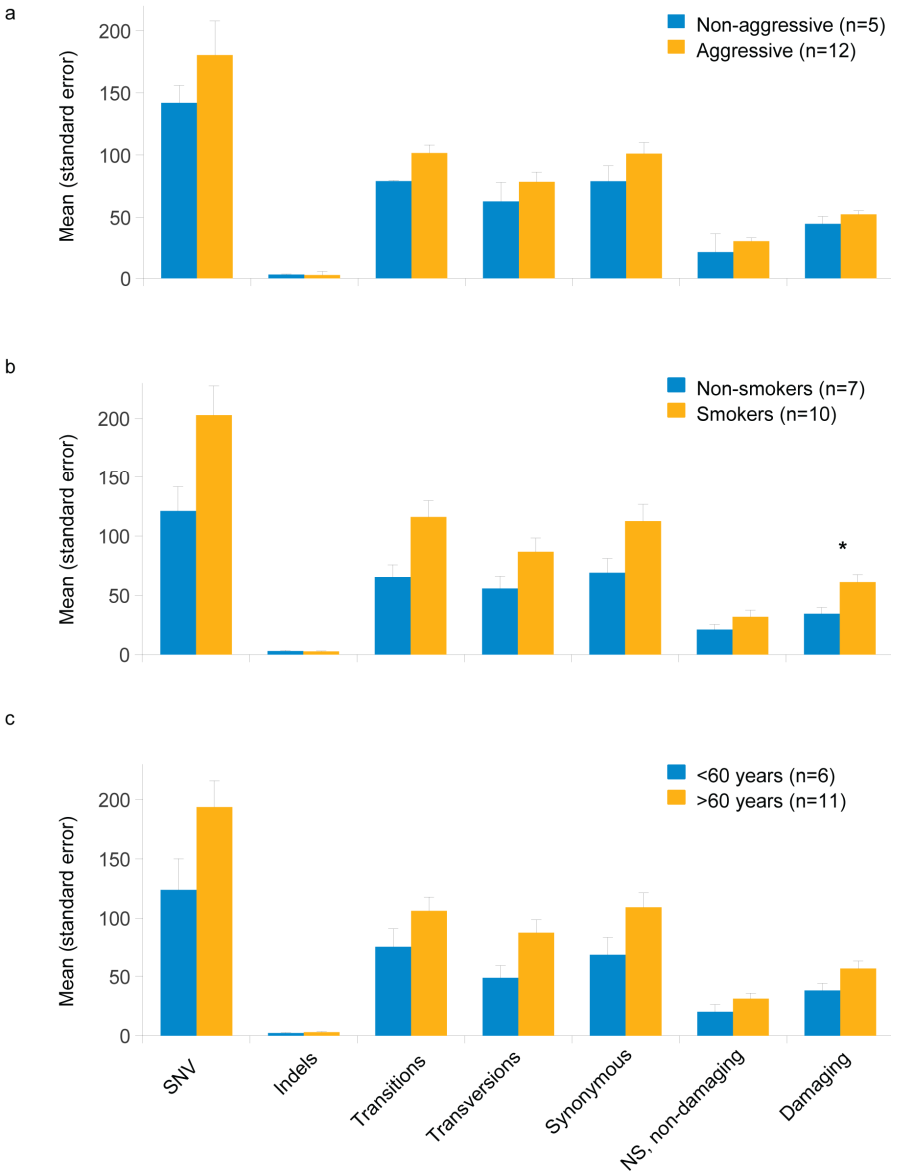

Figure 9. Number of mutations identified classified by variant type. (a) According to tumor aggressiveness. (b) According to smoking status. (c) According to patient age. None of the comparisons was statistically significant. *P=0.09.

1.3. HaloPlex targeted resequencing

To extend the analysis and identify genes that are frequently mutated in bladder cancer, we selected 110 genes either found to be recurrently mutated in the exome sequencing study or belonging to the same pathways as a recurrently mutated gene (Supp. Table 4). We used the HaloPlex Target Enrichment System followed by multiplex sequencing on 60 additional tumors, 51 of which were NMI (see Supp. Table 1 for patient characteristics, and Supp. Table 7 for depth of

coverage information). Bioinformatics analysis identified 260 somatic variants. Of them, 200 were predicted to be relevant and 143 predicted to be damaging (see Materials and methods). Sanger sequencing verified 73/95 SNVs analyzed (76.13%); 72 (98.6%) of the verified alterations were found to be somatic.

1.4. The mutational landscape of UBC in this study

Using the variants called by the discovery and prevalence screens, we compiled a list of genes found to harbor relevant mutations in ≥3 tumors. We used microarray expression data from 43 tumors covering the UBC spectrum to select only genes found to be expressed in >30% of UBC. The results are shown in Figure 10 and summarized in Table 5. As expected, well-established players in bladder tumor development, such as *FGFR3, PIK3CA, RB1*, and *TP53*, were among the recurrently mutated genes.

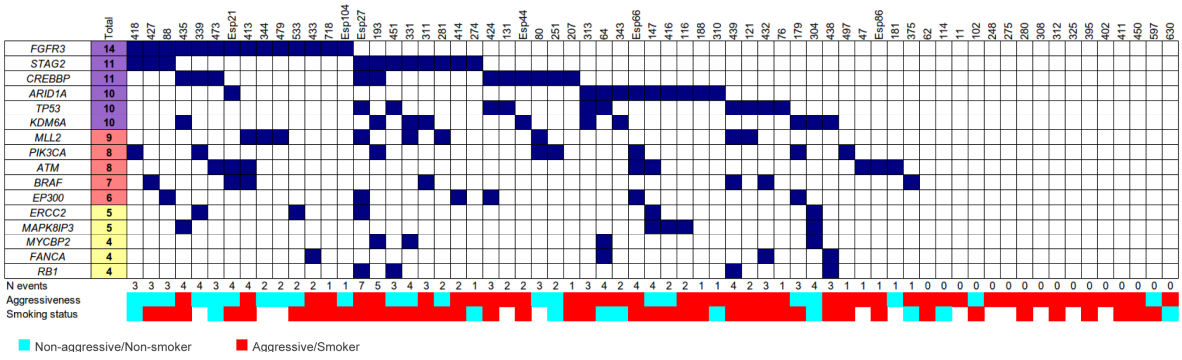

Figure 10. Cases harboring mutations in recurrently altered genes expressed in ≥30% of UBC in both the exome sequencing and the HaloPlex targeted resequencing. Each dark blue square represents one mutation. 22/77 (28.6%) of cases did not exhibit any mutations in these loci. For tumor aggressiveness, non-aggressive cases are depicted in blue, and aggressive tumors in red. For smoking status, non-smokers are depicted in blue and smokers in red. Cases where information was not available are depicted in white.

1.5. The chromatin remodeling, apoptosis, DNA repair and DNA damage response, and cell cycle pathways are frequently altered in UBC

To gain an understanding of the biological processes altered in UBC, we classified all relevant SNVs identified in the exome sequencing study according to their damaging potential and recurrence (see Materials and methods) and performed pathway analysis according to KEGG pathways and GO (Gene Ontology biological process) components. We found the chromatin modification, DNA repair and DNA damage response, apoptosis, and cell cycle pathways to be recurrently altered in UBC (false discovery rate (FDR) < 0.1). As shown in Supplementary Table 8, all pathways were significantly altered when analyzing only SNVs predicted to be damaging, and the chromatin modification, apoptosis, and DNA repair and DNA damage response pathways, but

not the cell cycle pathway, were still significantly altered when taking into account only recurrent SNVs, as well as recurrent mutations predicted to be damaging.

Table 5. **Expressed genes frequently mutated in UBC**. *P-values calculated based on mutations identified in the exome sequencing screen. **P-values calculated comparing non-aggressive and aggressive tumors. NA: Not available.

GENE	**Number of mutations in exome screen (n=17)**	***P*-value***	**Number of mutations in prevalence screen (n=60)**	**Number of mutations in all tumors (n=77)**	**Number of non-aggressive mutant cases (n=29)**	**Number of aggresssive mutant cases (n=47)**	***P*-value****
ARID1A	7	0.0001	3	10	3	7	0.732
STAG2	3	0.019	9	12	6	5	0.315
KDM6A	4	0.019	6	10	3	7	0.732
PDZD2	3	0.019	0	3	0	2	0.521
MYCBP2	3	0.061	2	5	2	2	0.999
LPHN3	3	0.096	0	3	0	2	0.521
CREBBP	2	0.098	9	11	4	7	1
EP300	2	0.098	5	7	3	4	1
ATM	3	0.138	6	9	4	4	0.702
TP53	3	0.2117	8	11	2	9	0.188
RREB1	3	0.237	0	3	1	1	1
PIK3CA	6	0.239	4	10	5	4	0.289
WHSC1L1	2	0.241	1	3	1	2	1
MYO5B	3	0.430	0	3	0	2	0.521
MLL2	2	0.636	13	15	6	5	0.315
FGFR3	2	0.659	12	14	10	4	0.011
TEX15	3	0.778	0	3	0	2	0.521
BRAF	1	NA	6	7	2	5	0.701
ERCC2	0	NA	8	8	5	1	0.040
MAPK8IP3	1	NA	4	5	3	2	0.363
MLL	1	NA	5	6	3	2	0.363
NUP93	1	NA	4	5	1	3	1
STAG1	0	NA	5	5	0	3	0.282
RB1	1	NA	3	4	1	3	1
FANCA	0	NA	4	4	0	3	0.282
MLL3	1	NA	4	5	2	3	1
NOTCH1	0	NA	4	4	0	3	0.282
ASXL2	3	NA	1	4	1	1	1

Within the chromatin remodeling pathway, we found mutations in genes not previously identified as being mutated in bladder cancer (*ASXL2*, *BPTF*, and *MLL2*), as well as in genes recently reported to be altered in UBC (*ARID1A*, *CREBBP*, *EP300*, *KDM6*, *MLL*, and *MLL3*) (Gui et

al. 2011). Interestingly, we also found mutations in several genes involved in DNA repair, including *ATM*, *ERCC2*, and *FANCA*; this group of genes had not been previously reported as being somatically mutated in bladder cancer. As of the cell cycle pathway, we identified mutations in several components of the cohesin complex, including *STAG1*, *SMC1A*, *SMC1*β, and *STAG2*, although only the latter was mutated in more than 3 cases.

2. *ARID1A*

2.1. ARID1A truncating mutations are frequent in aggressive UBC

The exome discovery and HaloPlex prevalence screens detected 10 mutations in *ARID1A* (Table 6). To expand these findings, we performed targeted exon resequencing of exons 2-20 of *ARID1A* in 48 additional primary tumors and in 5 UBC cell lines. We also included 1 tumor that had been analyzed by exome sequencing and 3 cases that were part of the HaloPlex prevalence screen (total n=52). Details on the average depth of reads per exon for the *ARID1A* sequencing study are shown in Table 7. Figure 11 shows sequencing breadth and depth for individual tumor and cell line samples.

Table 6. **Relevant mutations in *ARID1A* identified in the discovery and prevalence screens.** [1]Identified in the exome sequencing study. [2]Identified in the HaloPlex prevalence screen. [3]Identified in the *ARID1A* mutational analysis.

Sample	Stage/Grade	Amino acid substitution	Affected exon
064	T1G2	[1,2]P459A	3
116	T1G2	[1]L1014 frameshift	11
188	T1G3	[1,2]S614*	4
310	T1G3	[1]K1745 frameshift	20
313	T1G3	[1]G1317 frameshift	16
343	T1G3	[1]Q393*	2
Esp66	T1G3	[1,3]Q403*	2
416	TaG1	[2]G660V	5
147	TaG1	[2]L1076I	12
Esp21	TaG1	[2,3]N2066D	20
Esp69	T3G3	[3]S769*	7
800	T1G3	[3]C2052*	20
803	T1G3	[3]S571L	3
		[3]S614*	4
559	T1G3	[3]Q393*	2
VMCUB-3	Cell line	[3]E1733*	20
		[3]D1738N	
		[3]Q2210H	
		[3]L1922L	

Considering the alterations present in at least 10% of the reads for the mutated nucleotide position, we found 6 tumors and one cell line harboring a total of 11 different SNVs (Table 6). We did not find any correlation between sequencing depth and the frequency at which variant alleles were detected (Figure 12a). All of the mutations identified in the *ARID1A* mutational study were independently confirmed with Sanger sequencing (Figure 13). Additionally, we found 6 mutations predicted to be damaging at a frequency <10% that were not validated by Sanger sequencing, possibly due to the lower sensitivity of this technique (Figure 12b).

Table 7. **Summary of reads per exon in the *ARID1A* resequencing study.** Exon starting and ending positions are shown, along with the exon length (base pairs) and the average sequencing depth.

ARID1A exon	Start	End	Length	Average depth
E1	1001	2510	1509	0
E2	34621	34833	212	9587
E3	36122	36574	452	18256
E4	37646	37762	116	6887
E5	65826	66066	240	8571
E6	66354	66443	89	8256
E7	67122	67289	167	6007
E8	67943	68255	312	10586
E9	71191	71336	145	10867
E10	71427	71536	109	9603
E11	72760	72969	209	11690
E12	76089	76296	207	9029
E13	77470	77602	132	5867
E14	77782	77957	175	8826
E15	78316	78466	150	4283
E16	78550	78687	137	3683
E17	78772	78868	96	11134
E18	79950	80190	240	16486
E19	80547	80677	130	17715
E20	83993	87080	3087	96890

Integrating the results from the targeted exon resequencing screen and those from the exome discovery and HaloPlex prevalence screens, we identified a total of 15 damaging mutations in 14/125 tumors (11.2%) (Table 6). There were 7 nonsense, 5 missense, and 3 frameshift mutations. One tumor presented 2 mutations, 1 truncating and 1 missense. All 10 tumors harboring nonsense (n=7) or frameshift (n=3) mutations belonged to the aggressive category. Three non-aggressive tumors presented 1 missense mutation each, and 2 aggressive tumors also harbored missense SNVs, although 1 of them also exhibited a mutation leading to a premature stop codon

(Table 6). Overall, we found truncating mutations in 0/43 non-aggressive and 7/82 aggressive tumors (P = 0.048).

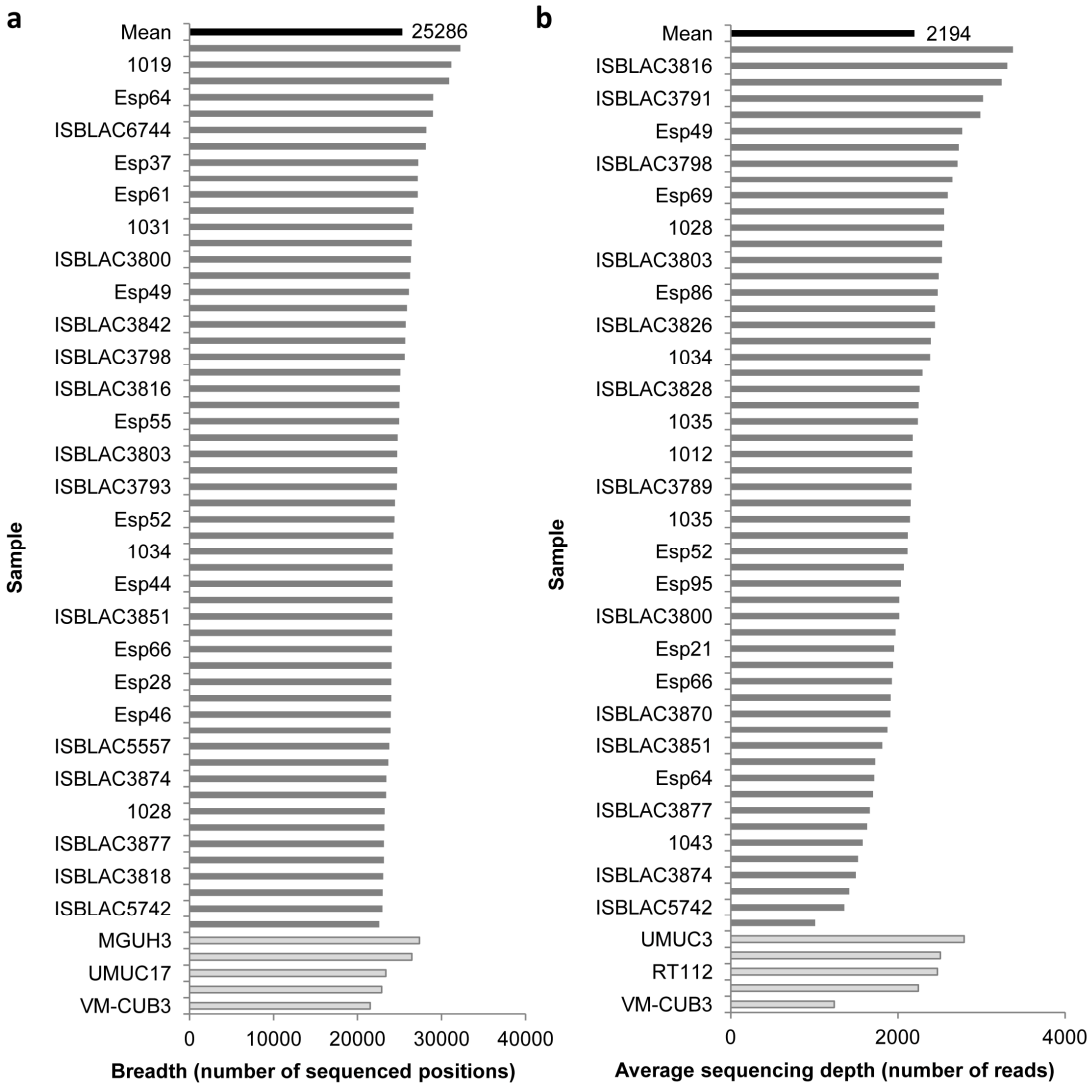

Figure 11. Data metrics for *ARID1A* resequencing. (a) Average sequencing breadth (reads per exon) per case. (b) Average sequencing depth (reads per exon) per case.

An analysis of the *ARID1A* mutations reported by 7 independent studies in the course of this work in UBC revealed 119 relevant mutations in 556 tumors published (21.4% overall mutation rate) (Gui et al. 2011; Guo et al. 2013; Cazier et al. 2014; CGARN 2014a; Kim et al. 2014; Nordentoft et al. 2014; Ross et al. 2014). Considering together the information from this study and all published work, UBC harbors *ARID1A* relevant alterations in 14.4% of NMI and 23.6% of MI cases (P=0.02) (Table 8).

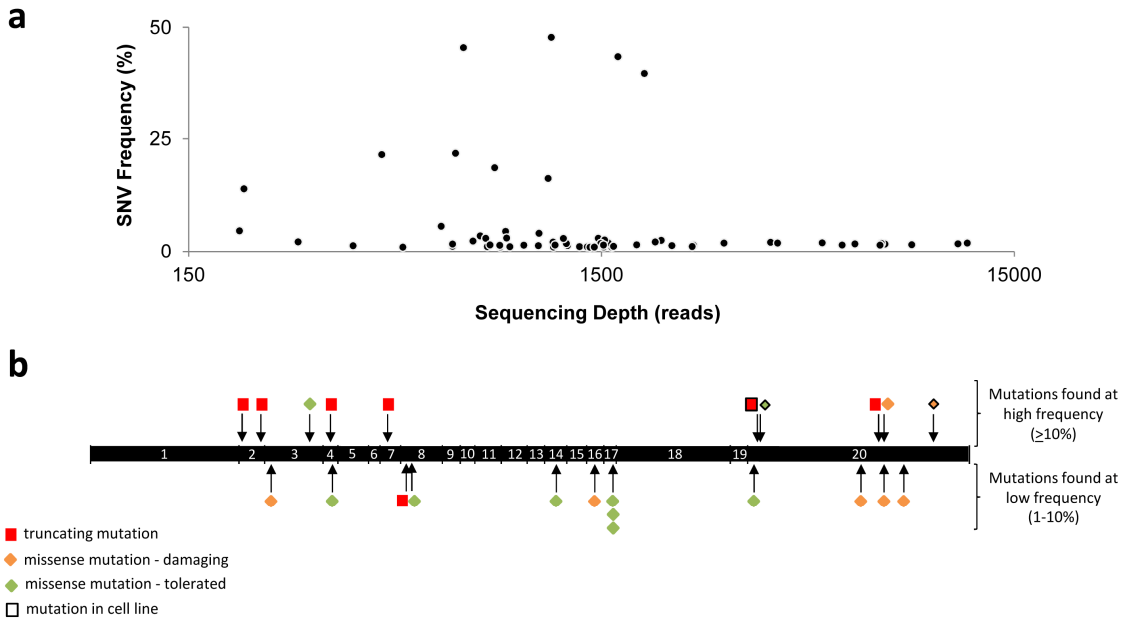

Figure 12. SNV frequency and type found in the *ARID1A* targeted exon resequencing study. (a) SNV frequency vs. sequencing depth. (b) Variant location and SIFT-predicted effect for alterations detected at high (top) and low (bottom) frequencies.

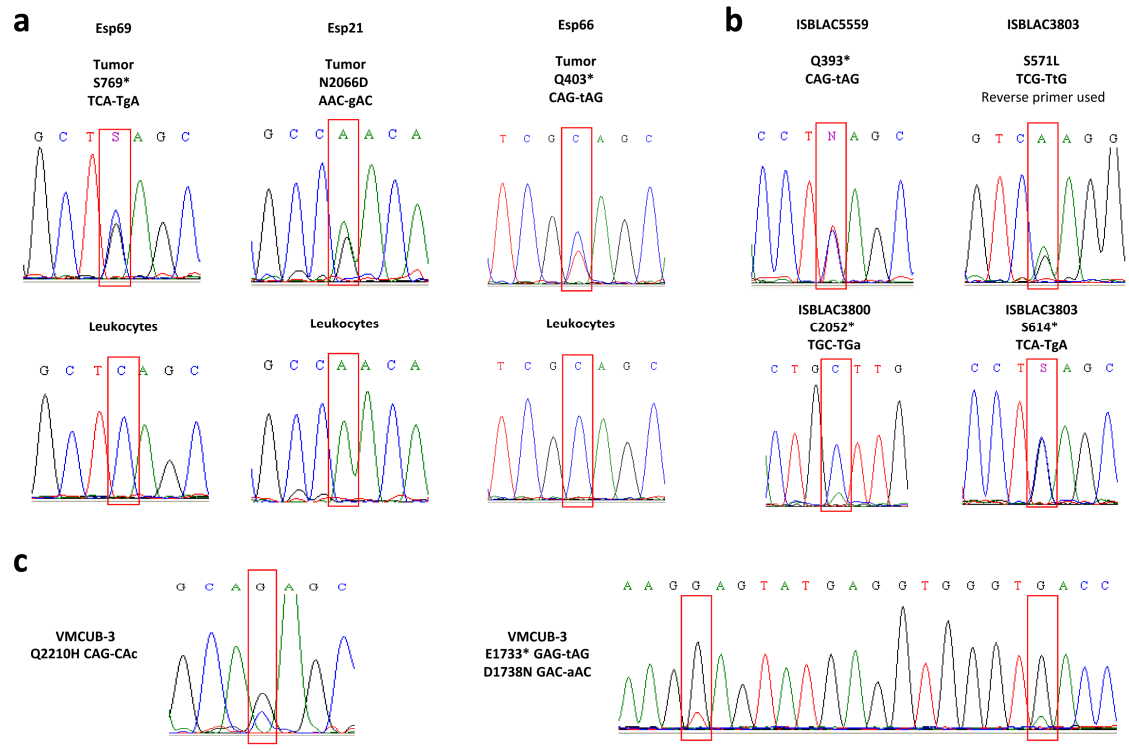

Figure 13. *ARID1A* mutations in UBC. All mutations identified at a frequency ≥ 10% were verified by Sanger sequencing of independent PCR products. Both the nucleotide change and the predicted effect on the amino acid sequence are shown. * denotes a truncating mutation. (a) Mutations found in three different tumors and their corresponding normal leukocyte DNA counterparts. (b) Mutations found in four additional tumors. (c) Three different mutations were identified in the VMCUB-3 UBC cell line.

Table 8. *ARID1A* mutations in all UBC studies.

Study	Study type	Total cases (mutated)	NMI-LR cases (mutated)	NMI-HR cases (mutated)	MI cases (mutated)
This study	WES + targeted	125 (14)	58 (5)	50 (8)	15 (1)
Gui et al. 2011 + Guo et al. 2013	WES + targeted	105 (16)	36 (4)	3 (1)	66 (11)
CGARN 2014a	WES	130 (39)	0 (0)	0 (0)	130 (39)
Nordentoft et al. 2014	WES	38 (10)	22 (8)	9 (0)	7 (2)
Cazier et al. 2014	WGS	14 (2)	4 (0)	5 (1)	5 (1)
Ross et al. 2014	targeted	35 (7)	0 (0)	0 (0)	35 (7)
Kim et al. 2014	targeted	109 (31)	NA	NA	NA
	Total	556 (119) = 21.4%	120 (17) = 14.2%	67 (10) = 14.9%	258 (61) = 23.6%

Regarding cell lines, RT112, MGH-U3, UM-UC-3, and UM-UC-17 were found to be wild type for *ARID1A*. However, VM-CUB3 cells harbored three *ARID1A* mutations: 1 truncating, 2 missense, and 1 silent mutation (Fig. 13, Table 6). RT112 cells are *TP53* wild type (Rieger et al. 1995), and express high levels of wild type *FGFR3*, all features reminiscent of low grade NMIUBC. On the other hand, VMCUB-3 cells harbor mutant *TP53* (Rieger et al. 1995).

2.2. Loss of ARID1A protein expression is associated with an aggressive phenotype

Truncating mutations in *ARID1A* have been shown to associate with loss of protein expression in gynecological malignancies (Wiegan et al. 2010; Guan et al. 2011a). We assessed *ARID1A* protein and mRNA levels in a panel of UBC cell lines (Fig 14a). We found several cell lines, including VM-CUB3, that did not express the protein, but there was not a consistent correlation between mRNA and protein levels.

We also found loss of ARID1A expression by IHC in a tumor from the exome screen harboring a truncating mutation (Figure 13a, 14b). Taking this into account, we performed an IHC analysis in a panel of 84 tumors, using ARID1A loss of expression as a surrogate for pathogenic mutations. ARID1A expression was lost significantly more often in MI tumors ($\geq$T2) (average histoscore 83.1 $\pm$ 91.1) than in high risk NMI (TaG3, T1G2, T1G3) (155.8 $\pm$ 110.1) and low risk NMI (TaG1, TaG2) (206.4 $\pm$ 80.2) tumors (ANOVA P= 9.9×10^{-6}; KW P=3×10^{-5}) (Supp. Table 9, Fig. 15a, top panel).

To place these findings in the context of the known pathways of UBC tumor progression, we compared *ARID1A* and *FGFR3* mutational status in 50 tumors for which both data were available.

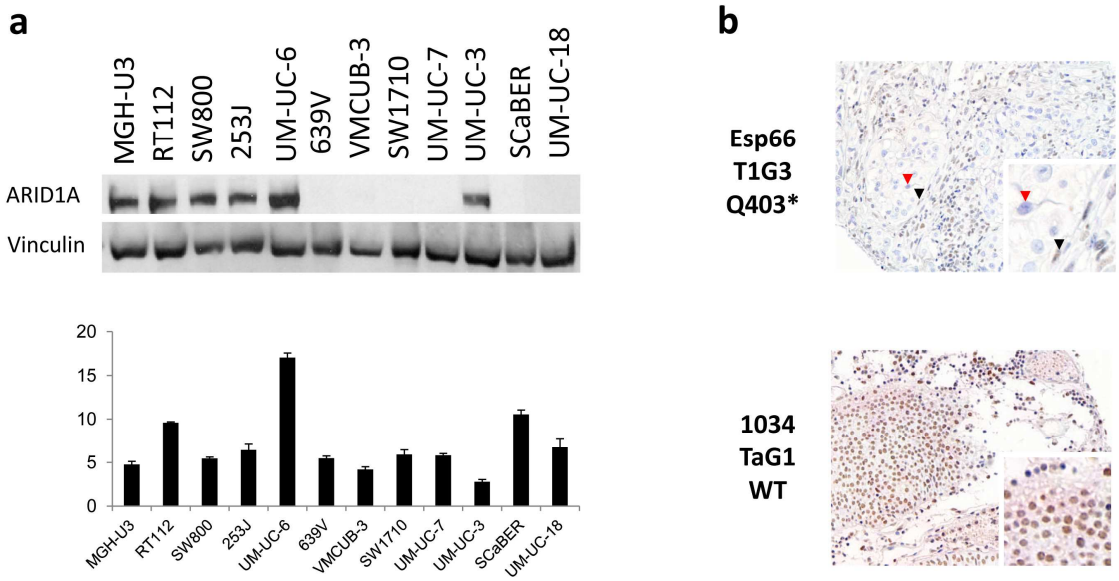

Figure 14. *ARID1A* expression in UBC. (a) ARID1A protein (top) and mRNA (bottom) expression in a panel of UBC cell lines. (b) An aggressive tumor with a truncating mutation in *ARID1A* losses protein expression (top), as opposed to a non-aggressive case with a wild type *ARID1A* sequence (bottom). The red arrowheads indicate tumor cells with loss of ARID1A; the black arrowheads point to stromal cells with positive ARID1A staining.

The 5 tumors with truncating *ARID1A* mutations were all wild type for *FGFR3* (p = 0.056, Fisher's exact test). This suggests that the two genes might be involved in different genetic pathways. However, the low number of tumors included in the analysis did not allow us to draw definite conclusions. Therefore, we assessed FGFR3 IHC expression, finding that histoscores were higher in low risk NMI (73.5 $\pm$ 67.2) than in MI tumors (47.8 $\pm$ 75.5) (ANOVA P= 0.038, KW P= 0.026) and significantly correlated with ARID1A expression (Spearman correlation P= 0.03) (Supp. Table 9, Fig. 15a middle panel). This indicates that alterations in *ARID1A* and *FGFR3* occur independently and tend to be mutually exclusive in MIUBC. To further explore the relationship between ARID1A and other UBC markers, we used IHC as a surrogate for the mutational status of *TP53*, another well-established molecular player in the development of UBC. The majority of *TP53* mutations result in its nuclear accumulation. Accordingly, IHC scores for nuclear p53 were higher as tumor stage and grade increased (16.2$\pm$17.5 vs. 82.5$\pm$126.1 and 133.8$\pm$128.2) (ANOVA P= 0.05, KW P= 0.32) but no significant correlation was found when comparing p53 scores with ARID1A expression (1 vs. -0.1, P =0.30) (Supp. Table 9, Fig. 15a bottom panel).

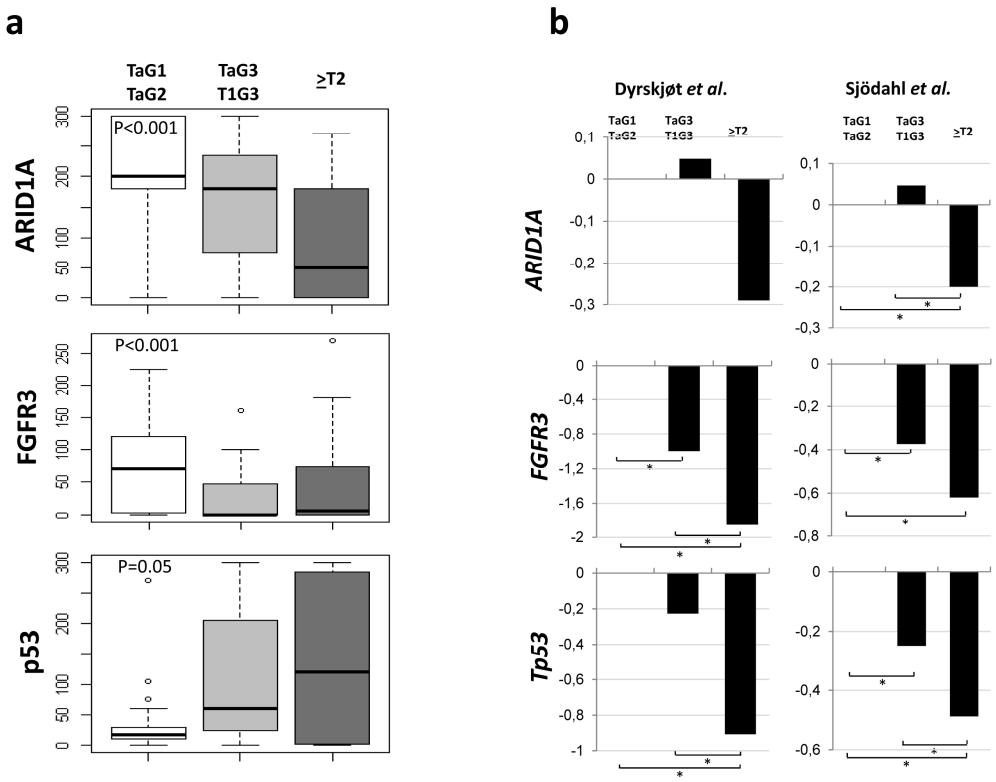

Figure 15. Loss of ARID1A protein expression associates with aggressive UBC. Tumors were classified into low grade NMI (TaG1, TaG2), high grade NMI (TaG3, T1G3), and MI (≥T2). (a) ARID1A staining is significantly lower in more aggressive tumors. The same pattern is observed for FGFR3, whereas p53 immunohistochemical score is higher in more advanced, aggressive tumors. (b) Differential expression of *ARID1A*, *FGFR3* and *TP53* mRNA in two independent UBC microarray studies. Levels of the three genes are significantly lower in more aggressive tumors. *indicates a False Discovery Rate-adjusted P-value <0.05.

We next set out to determine whether *ARID1A, FGFR3,* and *TP53* are differentially expressed at the RNA level in the low risk NMI, high risk NMI, and MI UBC tumor subgroups. We analyzed 2 public expression datasets (Dyrskjot et al. 2003, Sjodahl et al. 2012). In both series, *ARID1A, TP53*, and *FGFR3* mRNA levels were significantly lower in MI tumors (Fig 15 b). This further supports our finding that *FGFR3* mutations are less frequent in aggressive tumors and associate with *FGFR3* overexpression at the mRNA level, whereas *TP53* mutations, more common in aggressive cases, associate with decreased mRNA expression.

2.3. ARID1A loss is associated with worse patient outcome

To explore whether ARID1A expression is associated with patient outcome, we performed Kaplan-Meier analysis on recurrence, progression, and mortality data (Fig. 16, Supp. Table 10). Patients whose tumors had a IHC score <180 presented a significantly lower rate of tumor

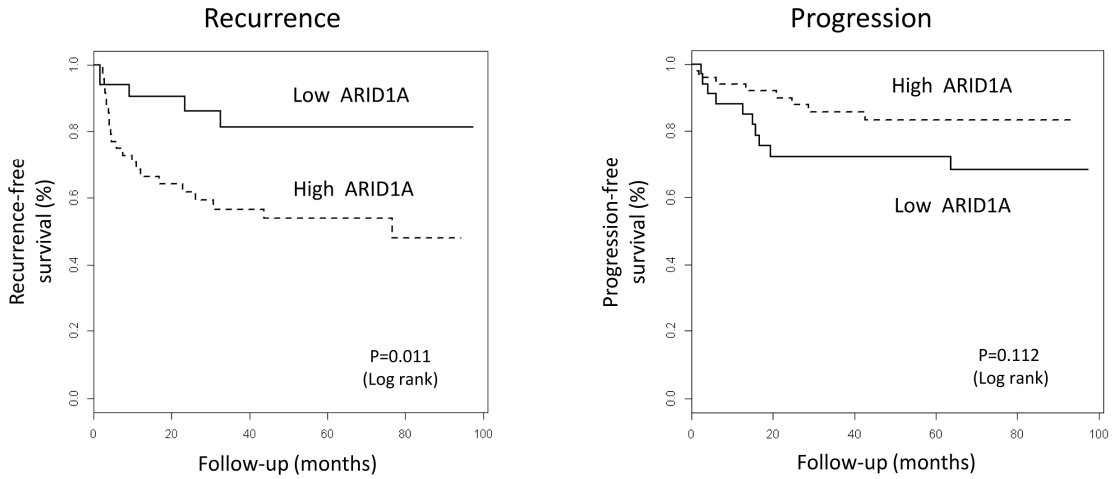

Figure 16. Low ARID1A protein levels associate with a poor outcome. ARID1A IHC score was calculated for 84 patients. The significantly lower risk of recurrence (left) and the increased risk of progression (right) are indicative of a more aggressive clinical course.

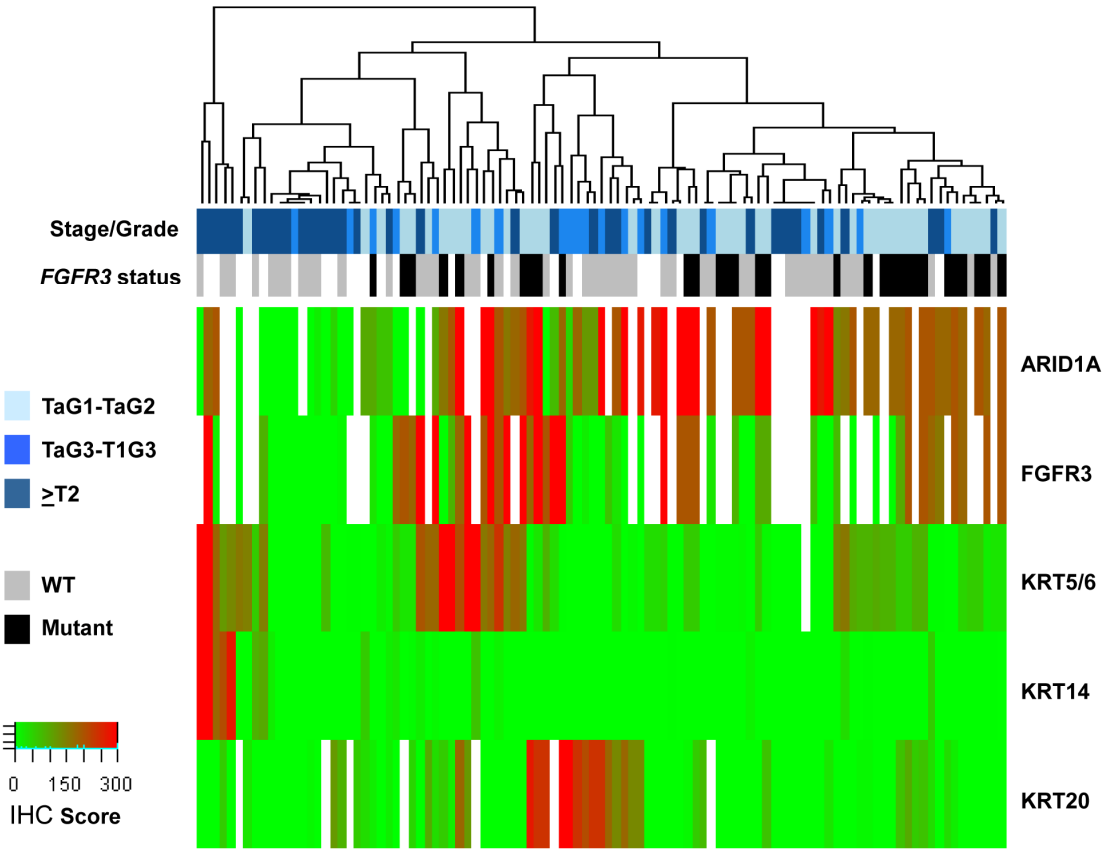

Figure 17. Unsupervised clustering of ARID1A, FGFR3, KRT5/6, KRT14, and KRT20 levels. IHC scores are showed in a green (lowest)-red (highest) color code. Bars below the dendogram include information about tumor stage and grade (tones of blue) and *FGFR3* mutational status (grey/black). White squares indicate information is not available.

recurrence (P=0.011) but exhibited a tendency to show an increased rate of tumor progression (P=0.112). This suggests that loss of ARID1A is associated with poor prognosis but further studies including a larger number of samples that allows for multivariable or stratified analyses will be required to adequately assess this relationship.

2.4. Association between ARID1A loss and differentiation markers

In UBC, altered differentiation has been shown to identify aggressive tumors (Volkmer 2012). To assess the relationship between ARID1A protein levels and urothelial differentiation, we performed unsupervised analysis of IHC scores for ARID1A, FGFR3 and KRT5/6, KRT14, and KRT20 (Fig. 17). This analysis confirmed that the most aggressive tumors (≥T2) expressing low levels of ARID1A tend to present low levels of wild type FGFR3, as well as low KRT20. A subset of those tumors also expresses high levels of KRT5/6, and KRT14, characteristic of basal urothelial cells.

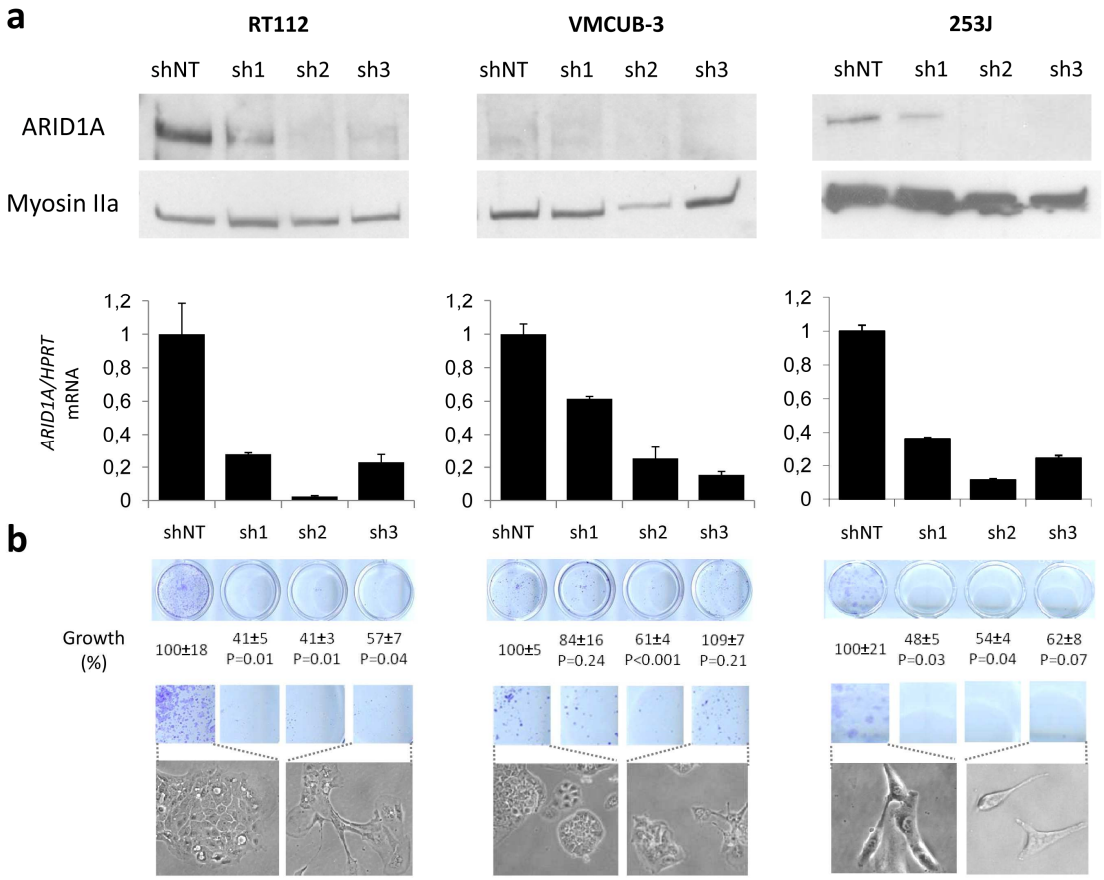

Figure 18. *Effects of knocking down ARID1A in the RT112, VMCUB-3, and 253J UBC cell lines*. (a) ARID1A knockdown with three different shRNAs was efficient at the protein (upper row) and mRNA (bottom row) levels. (b) ARID1A knockdown consistently and significantly associates with reduced colony formation in wild type RT112 and 253J but not in mutant VMCUB-3. Growth was quantified in triplicate experiments and is expressed in percentage ± Standard Error of the Mean. The bottom row shows representative morphological changes in cell lines interfered with control or *ARID1A*-targeting shRNAs.

2.5. ARID1A knockdown reduces cell viability in vitro

To address whether ARID1A plays a role in the control of cell proliferation, we knocked it down in wild type ARID1A-expressing UBC cells. Out of the 12 UBC lines assessed, 6 lacked detectable levels of protein by western blotting (Fig. 14a). In both RT112 and 253J lines, knocking down ARID1A with 3 different lentiviral shRNA constructs led to a significant reduction in colony formation ability, as well as to a flatter morphology, in 3 independent experiments (Fig. 18b). However, mutant VMCUB-3 cells did not exhibit consistent changes in colony formation ability upon ARID1A knockdown. The dramatic reduction in cell viability upon ARID1A knockdown did not permit performance of additional functional studies to assess cell differentiation, invasion or migration.

3. *STAG2*

3.1. Expression of several components of the cohesin complex is frequently lost in UBC

Given the fact that we found mutations in genes coding for several cohesin components, we analyzed IHC expression of STAG1, STAG2, SMC1, SMC3, PDS5B, and RAD21 in normal urothelium, finding all cell layers express these proteins (Fig. 19b). PDS5A was not analyzed due to the inadequate quality of the available antibodies. To get an overview of the expression of cohesin components in UBC, we used IHC to analyze 91 samples representative of the disease spectrum. We defined loss of expression as a histoscore $\leq$50 and we only considered cases where stromal cells exhibited clear positive staining. We observed a small fraction of tumors exhibiting loss of expression of all cohesin components assayed (4.3% for STAG1, 4.4% for SMC1, 5.2% for PDS5B, 5.4% for RAD21, and 5.7% for SMC3) (Fig. 19c). Interestingly, STAG2 expression was lost in 33.7% of cases. Unsupervised analysis of IHC scores revealed that most tumors harboring loss of expression of other cohesin subunits also exhibited decreased STAG2 expression (Fig. 20).

3.2. Meiotic cohesins are expressed in UBC

The exome screen described earlier also identified two mutations in *SMC1*β, a component that replaces SMC1A in meiotic cohesin. We found that *SMC1*β is expressed at the RNA level in 54% of UBCs. In addition, the transcripts coding for the meiotic cohesin components *STAG3* (which replaces STAG1/2) and *REC8* (taking the place of RAD21) were also expressed in 84% and 100% of UBCs respectively, as determined by Affymetrix microarray expression analyses (Fig. 21a, left panel). To explore the potential interest of these findings, we searched for antibodies of adequate specificity to assess expression in a panel of UBC cell lines. However, all the available antibodies tested failed a rigorous specificity test (Fig. 21a, right panel). A commercial antibody

detecting STAG3 has been reported to be specific using immunoblotting assays and a band with the mobility expected for STAG3 was found in lysates of 8 UBC cell lines (Fig. 21b). To further confirm the specificity of the antibodies for IHC assays, we used testis from *STAG3* wild-type and KO mice (kindly provided by Dr. Alberto Pendás) and found non-specific, non-nuclear staining in both (Fig. 21c). This prevented us from further pursuing his line of inquiry.

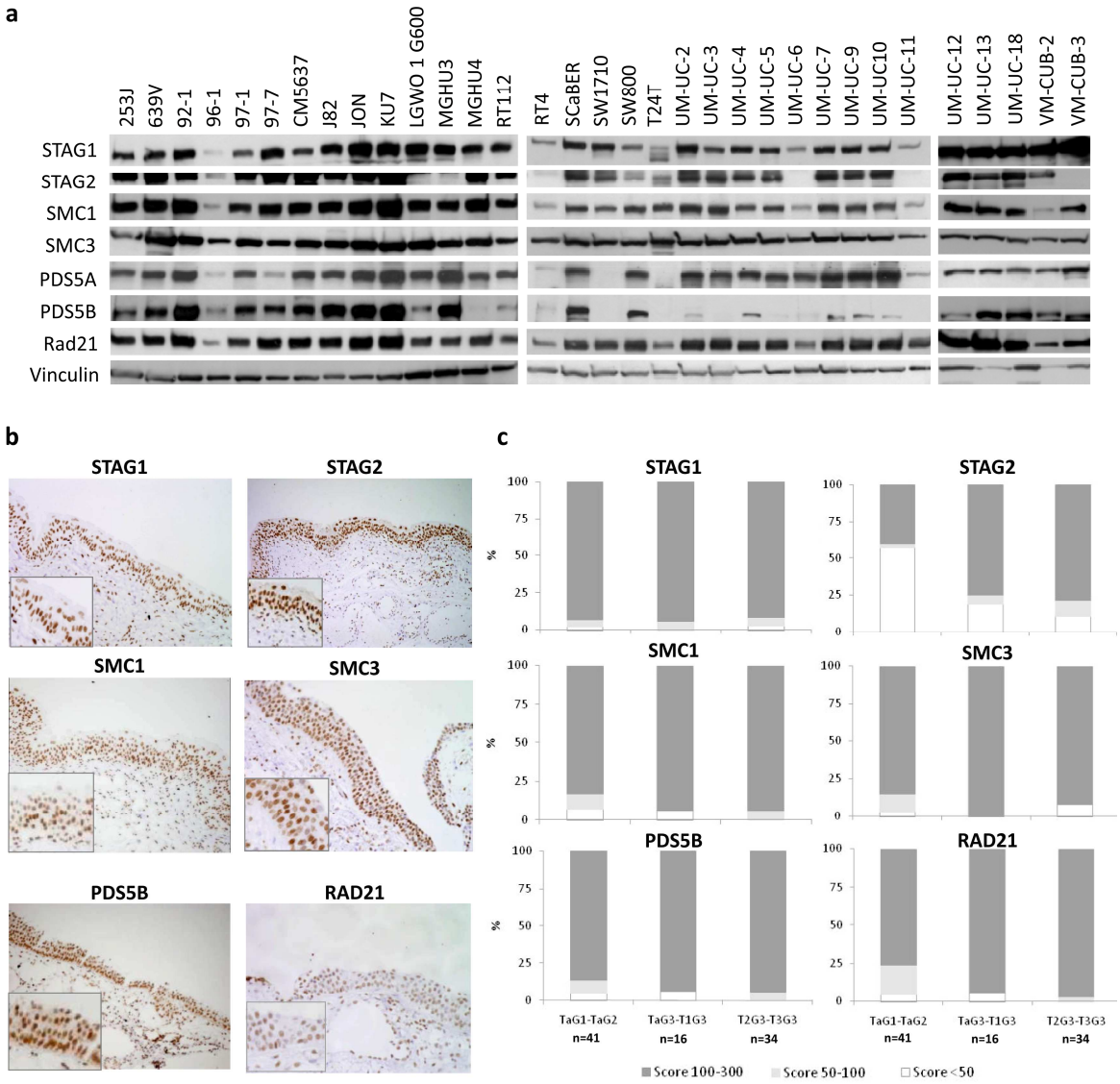

Figure 19. Expression of mitotic cohesin components in UBC cell lines, normal urothelium and UBC tumors. (a) Immunoblotting of different cohesin components in 33 different UBC cell lines. (b) Immunohistochemical expression of cohesins is detected in all layers of normal urothelium. (c) Immunohistochemical expression of cohesins in 91 UBC cases. Tumors were classified into low grade NMI (TaG1, TaG2), high grade NMI (TaG3, T1G3), and MI (≥T2). The percentage of tumors expressing high levels of STAG2 is represented in dark grey, those with reduced levels in light grey, and tumors with complete loss of expression are portrayed in white.

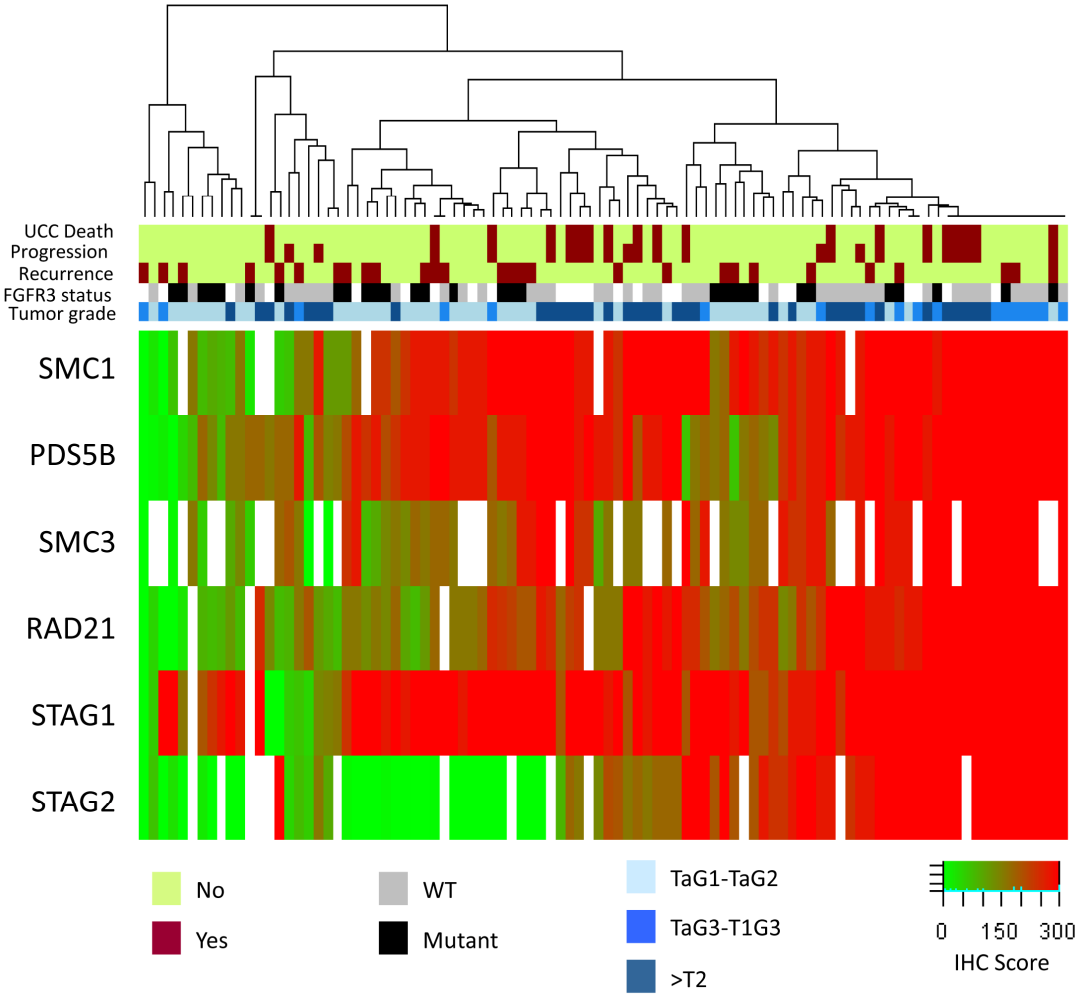

Figure 20. Non-hierarchical clustering of levels of six cohesion components. IHC scores are showed in a green (lowest)-red (highest) color code. Bars below the dendogram include information about prognosis (pistachio/bourbon), tumor stage and grade (tones of blue), and *FGFR3* mutational status (grey/black). White squares indicate information is not available.

3.3. STAG2 mutations and loss of expression are frequent events in UBC

Given the high frequency of loss of STAG2 expression and the fact that we found frequent mutations in our discovery and prevalence screens, we focused on STAG2. Overall, we identified 3 mutations predicted to be damaging in our exome sequencing study as well as 9 additional mutations predicted to be damaging in the HaloPlex resequencing analysis for a total of 12 damaging mutations in 77 tumors analyzed (15.6%) (Table 9). Of those, 5 were nonsense, 4 fell within intron-exon junctions, 2 were missense, and 1 was an indel. Mutations were distributed along the whole *STAG2* gene (Table 9, Fig. 22). We verified 9/11 mutations by DNA sequencing (Fig. 23). Damaging mutations were found both in non-aggressive (6/29, 20.7%) and in aggressive (5/47, 10.6%) tumors. A review of the reports concurrent and from other UBC sequencing studies

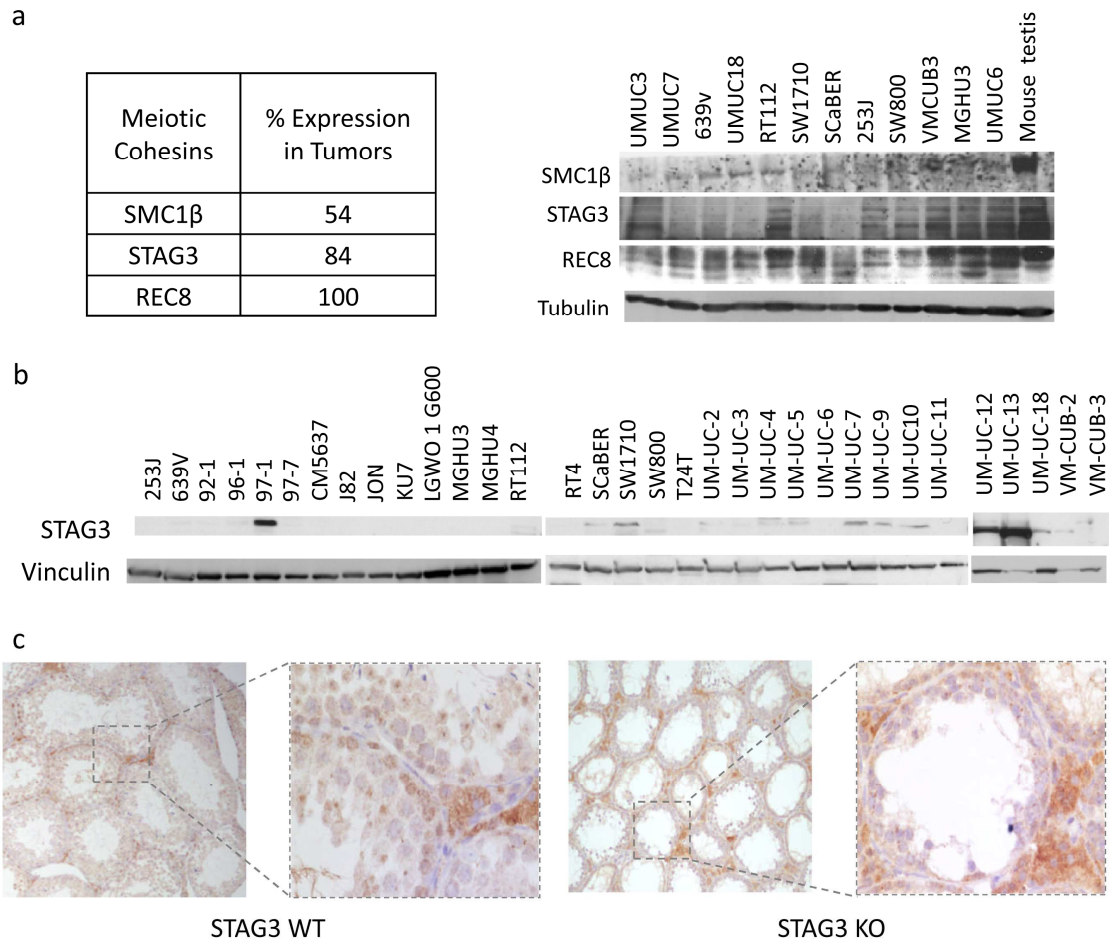

Figure 21. Meiotic cohesins are expressed in UBC. (a) The three meiotic-specific members of the cohesin complex are expressed in >50% of UBC. Homemade antibodies against SMC1β, STAG3, and REC8 (kindly provided by Dr. José Luis Barbero) showed multiple banding suggesting unspecific staining. (b) A commercial antibody targeting STAG3 shows mild to strong STAG3 expression in several UBC cell lines. (c) Immunohistochemical staining of testis from STAG3 wild type and KO mice.

(Gui et al. 2011; Guo et al. 2013; Solomon et al. 2013; Cazier et al. 2014; CGARN 2014a; Kim et al. 2014; Nordentoft 2014; Taylor et al. 2014) (Table 10, Fig. 22). In total, 17.3% of analyzed cases harbor a relevant *STAG2* mutation. Considering those cases for which adequate clinical and pathological information is available, *STAG2* relevant mutations are found in 29.3% of non-aggressive and 15.7% of aggressive cases (P< 0.0001) (Table 10).

Table 9. Relevant mutations in STAG2 identified in this study. * indicates a mutation introducing a premature stop codon.

Sample	Stage/Grade	Amino acid substitution	Affected exon	Study
88	TaG1	Exon Junction	7/8	Haloplex
193	T2G3	Exon Junction	24/25	Exome
274	T1G2	R216*	8	Exome/Haloplex
281	TaG2	R614*	20	Haloplex
311	TaG3	W485*	16	Haloplex
331	TaG1	Exon Junction	21/22	Exome
414	T1G3	D1253H	34	Haloplex
418	TaG1	Q573*	18	Exome
427	TaG1	Exon Junction	14/15	Haloplex
451	TaG2	R216*	8	Exome/Haloplex
Esp27	T2G3	D898N	27	Haloplex

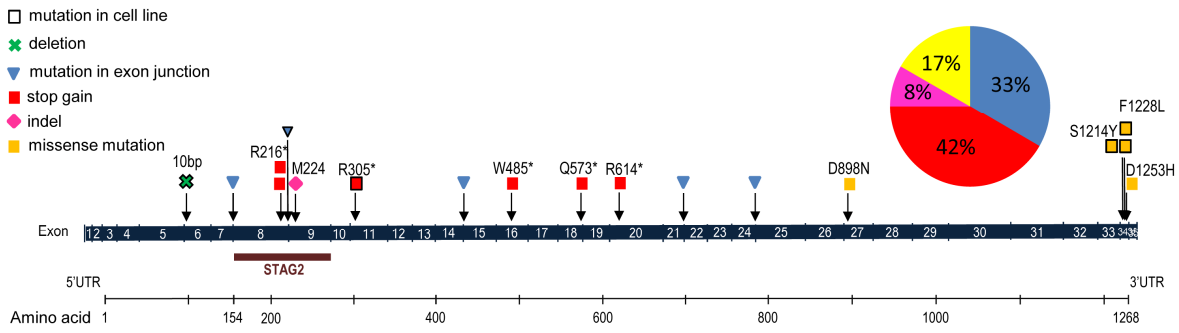

Figure 22. Summary of type and distribution of STAG2 nonsynonymous somatic mutations found in our discovery and prevalence screens.

Solomon et al. have reported, in non-epithelial tumors, that truncating mutations and deletions in *STAG2* lead to loss of protein expression (2011). We assessed STAG2 expression by IHC on 41 tumors for which mutational status had been analyzed and tissue sections were available. Out of 7 tumors with damaging mutations, 6 had lost expression, and 3/34 cases harboring wild-type STAG2 did not express the protein either (P=0.0001) (Fig. 23, Table 10).

Table 10. STAG2 mutations in all UBC studies.

Study	Study type	Total cases (mutated)	NMI-LR cases (mutated)	NMI-HR cases (mutated)	MI cases (mutated)
This study	WES + targeted	77 (11)	34(7)	33 (2)	9 (2)
Gui et al. 2011 + Guo et al. 2013	WES + targeted	155 (16)	36(8)	3 (0)	66 (4)
CGARN 2014a	WES	130 (14)	0 (0)	0 (0)	130 (14)
Solomon et al. 2013	targeted	111 (23)	24 (8)	29 (7)	58 (8)
Taylor et al. 2014	targeted	306 (78)	127 (43)	99 (26)	80 (9)
Nordentoft et al. 2014	WES	38 (7)	NA	NA	NA
Cazier et al. 2014	WGS + targeted	49 (4)	7(2)	2 (1)	5 (1)
Kim et al. 2014	targeted	109 (16)	NA	NA	NA
	Total	975 (169) = 17.3%	228 (68) = 29.8%	166 (36) = 21.7%	348 (38) = 10.9%

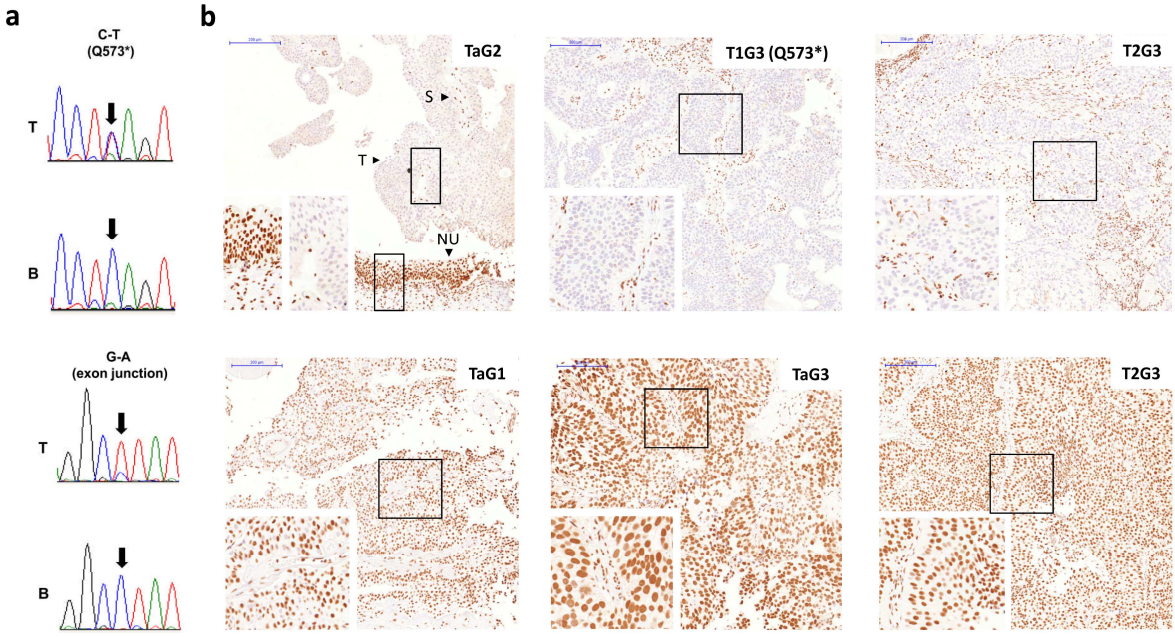

Figure 23. *STAG2* mutations and expression. (a) Sanger sequencing verification of a truncating (top) and exon-junction (bottom) mutation. T= tumor DNA, B= normal blood DNA. (b) Immunohistochemical analysis of STAG2 expression in UBCs of different stage and grade. Left column: low risk NMI tumors with and without loss of expression. T= tumor, S= stroma, NU= normal urothelium; the normal urothelium expresses STAG2. Middle column: high risk NMI tumors with and without loss of expression. Right column: MI tumors with and without loss of expression. STAG2 expression was found in the stroma of all tumor samples.

To determine whether other genetic mechanisms can lead to STAG2 inactivation in bladder cancer, we used data from SNP arrays corresponding to tumors included in our screens. STAG2 is the most frequently lost cohesin gene at the genetic level (Fig. 24a). We further analyzed genomic information from 18 males with UBC from whom we had available tumor tissue samples, finding 1 case (5%) in which a focal chromosome X genomic loss was responsible for loss of STAG2 expression (Fig. 24b).

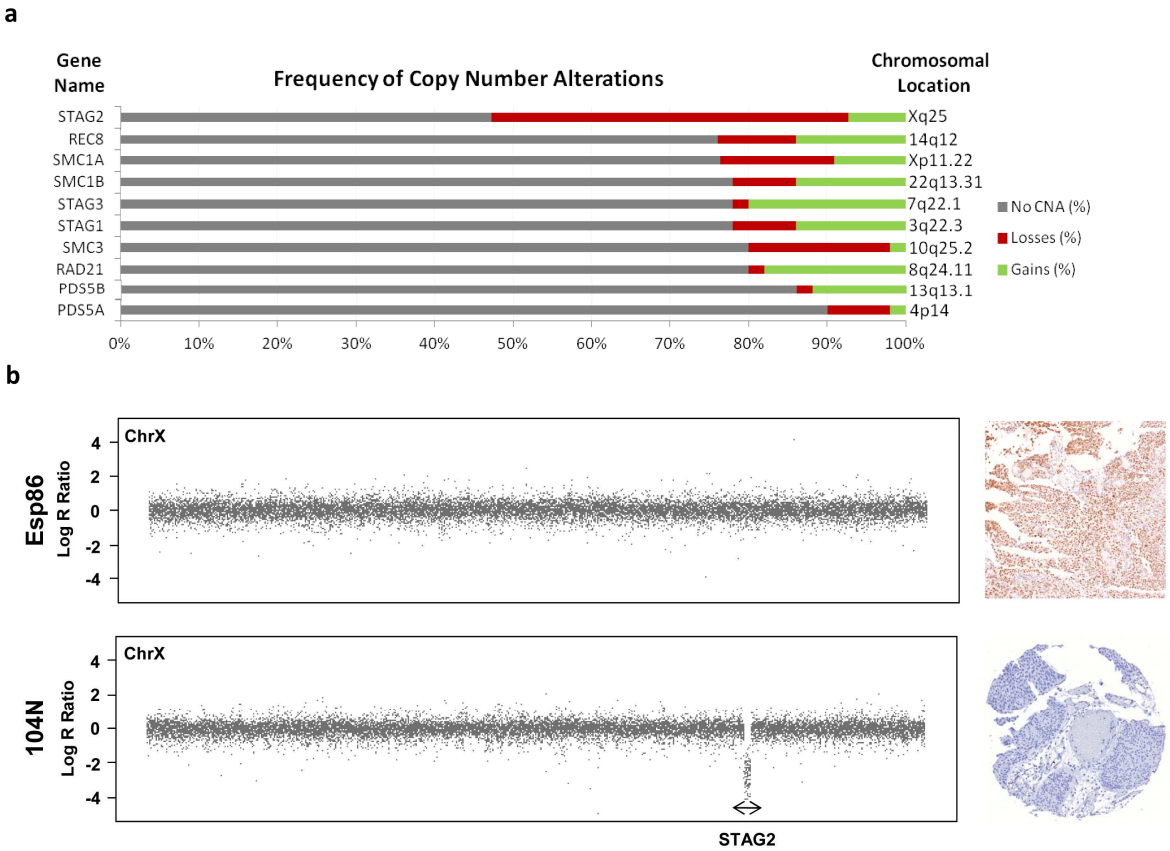

Figure 24. Genomic *STAG2* losses in males contribute to loss of protein expression in UBC. (a) Frequency of copy number alterations (CAN) in cohesin components in UBC. The bars represent percentage of tumors exhibiting losses (red), gains (green) or no alterations (grey) at the given chromosomal locations. (b) A male patient without genomic changes in the X chromosome expresses STAG2 (upper panel) but another one with a focal loss at the *STAG2* locus losses expression at the protein level (lower panel).

To confirm and extend these observations, we used IHC to analyze STAG2 expression in a panel of 671 incident UBC covering the full disease spectrum (Supp. Table 3). We observed loss of STAG2 expression in 197/671 tumors (29.3%) (Fig. 23). In a small fraction of cases (3%) STAG2 expression was focally lost, indicating that inactivation of STAG2 in these cases might be an early, but not initiating, event in carcinogenesis. Additionally, 11% of tumors exhibited STAG2

cytoplasmic localization (Fig. 25), suggesting that some alterations might hinder nuclear import of the protein or make it more prone to nuclear export.

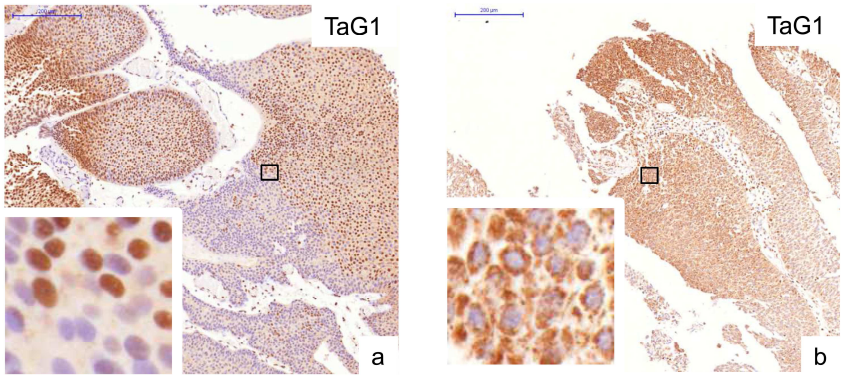

Figure 25. Abnormal STAG2 expression in UBC. (a) A non-aggressive tumor exhibits heterogeneous STAG2 expression. (b) A non-aggressive tumor shows cytoplasmic but not nuclear STAG2 expression. Scale bar: 200 µm.

3.4. Loss of STAG2 expression associates preferentially with non-aggressive characteristics

We analyzed the association between STAG2 expression and patient/tumor characteristics. STAG2 loss significantly associates with low stage ($P=5.7x10^{-15}$), low grade ($P=1.96x10^{-15}$), multicentricity (P=0.011), smaller tumor size (P=0.002), and lack of muscle invasion ($P=2.71x10^{-8}$) (Supp. Table 11, Fig. 26). We did not find any significant association between STAG2 loss and patient gender, age, geographical origin, or number of tumors (Supp. Table 11).

We then explored the relationship between STAG2 loss and other alterations characteristic of the NMI pathway of UBC development. In NMI tumors, loss of STAG2 was significantly associated with mutant *FGFR3* status (42.7% mutant vs. 27.2% wild type, P=0.001), lack of p53 overexpression (P=0.003), and a low Ki67 proliferative index (P=0.002) (Table 11).

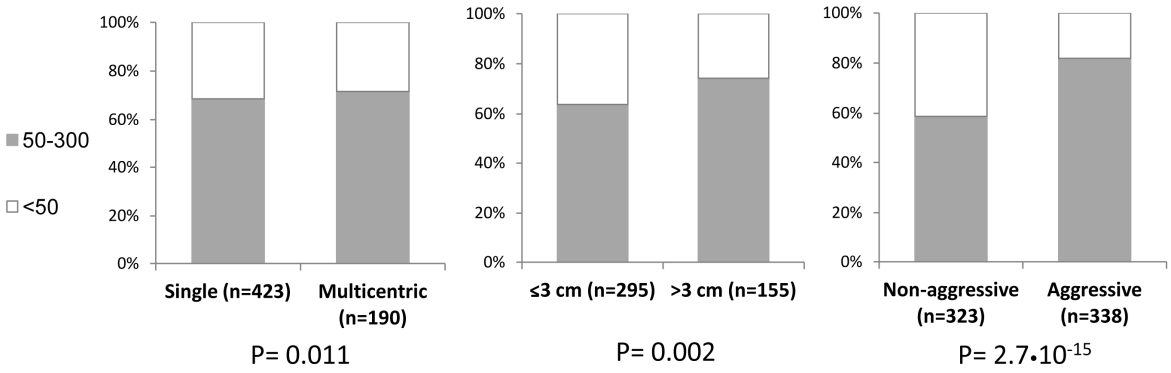

Figure 26. Loss of STAG2 expression significantly associates with features of low aggressivity. STAG2 expression is lost more frequently in tumors presenting single rather than multicentric locations (left panel), in smaller tumors (middle panel) and in those of non-aggressive stages and grades (right panel).

Table 11. **Association between STAG2 expression and other markers in NMIUBC.**

	N	STAG2 ≤50 N(%)	STAG2 > 50 N(%)	P-value *	OR (95% CI)
FGFR3 WT	169	46 (35.94)	123 (54.19)	**0.001**	0.47 (0.3-0.74)
FGFR3 Mut	186	82 (64.06)	104 (45.81)		
FGFR3 ≤30 **	224	71 (43.29)	153 (58.17)	**0.004**	0.55 (0.37-0.81)
FGFR3 >30	203	93 (56.71)	110 (41.83)		
Ki67 ≤15 $	361	150 (93.75)	211 (82.75)	**0.002**	3.13 (1.53-6.41)
Ki67 >15	54	10 (6.25)	44 (17.25)		
p53 ≤100 **	354	145 (89.51)	209 (77.12)	**0.002**	2.53 (1.42-4.51)
p53 >100	79	17 (10.49)	62 (22.88)		

* P-value of the chi-square test

** Histoscore; range 0-300

$ Percentage positive cells

Table 12. **Association between STAG2 expression and other markers in low risk NMIUBC.**

	N	STAG2 ≤50 N(%)	STAG2 > 50 N(%)	P-value *	OR (95% CI)
FGFR3 WT	86	27 (27.55)	59 (40.14)	0.059	0.57 (0.33-0.99)
FGFR3 Mut	159	71 (72.45)	88 (59.86)		
FGFR3 ≤30 **	138	50 (37.88)	88 (47.06)	0.13	0.69 (0.44-1.08)
FGFR3 >30	181	82 (62.12)	99 (52.94)		
Ki67 ≤15 $	286	125 (94.7)	161 (88.95)	0.11	2.22 (0.91-5.41)
Ki67 >15	27	7 (5.3)	20 (11.05)		
p53 ≤100 **	292	126 (96.92)	166 (88.3)	**0.011**	4.17 (1.4-12.42)
p53 >100	26	4 (3.08)	22 (11.7)		

* P-value of the chi-square test

** Histoscore; range 0-300

$ Percentage positive cells

Assessing exclusively the low-risk NMIUBC (non-aggressive) subgroup, STAG2 loss showed a borderline association with mutant FGFR3 (P=0.059) and low p53 expression (P=0.011) (Table 12),

whereas in high risk NMIUBC, STAG2 loss of expression was significantly associated with increased FGFR3 expression (P=0.037), and a low Ki67 proliferative index (P=0.049) (Table 13). However, neither NMI nor MI tumors showed associations between STAG2 loss and p53 expression (Tables 13, 14).

Table 13. **Association between STAG2 expression and other markers in high risk NMIUBC.**

	N	STAG2 ≤50 N(%)	STAG2 > 50 N(%)	P-value *	OR (95% CI)
FGFR3 WT	83	19 (63.33)	64 (80)	0.12	0.43 (0.17-1.09)
FGFR3 Mut	27	11 (36.67)	16 (20)		
FGFR3 ≤30 **	86	21 (65.62)	65 (85.53)	**0.037**	0.32 (0.12-0.85)
FGFR3 >30	22	11 (34.38)	11 (14.47)		
Ki67 ≤15 $	75	25 (89.29)	50 (67.57)	**0.049**	4 (1.1-14.57)
Ki67 >15	27	3 (10.71)	24 (32.43)		
p53 ≤100 **	62	19 (59.38)	43 (51.81)	0.6	1.36 (0.59-3.11)
p53 >100	53	13 (40.62)	40 (48.19)		

* P-value of the chi-square test

** Histoscore; range 0-300

$ Percentage positive cells

Table 14. Association between STAG2 and p53 expression in MIUBC.

	N	STAG2 ≤50 N(%)	STAG2 > 50 N(%)	P-value *	OR (95% CI)
p53 ≤100 **	85	9 (37.5)	76 (48.1)	0.45	0.65 (0.27-1.57)
p53 >100	97	15 (62.5)	82 (51.9)		

* P-value of the chi-square test

** Histoscore; range 0-300

STAG2 loss was more frequent in non-aggressive tumors, which are more differentiated, raising the possibility that loss might reflect urothelial cell differentiation. We confirmed STAG2 is expressed in all cell layers of normal urothelium (Fig. 27a), including the umbrella cells identified using antibodies detecting UPK3B and KRT20 (Fig.27b). This indicates that STAG2 expression is independent of urothelial cell maturation.

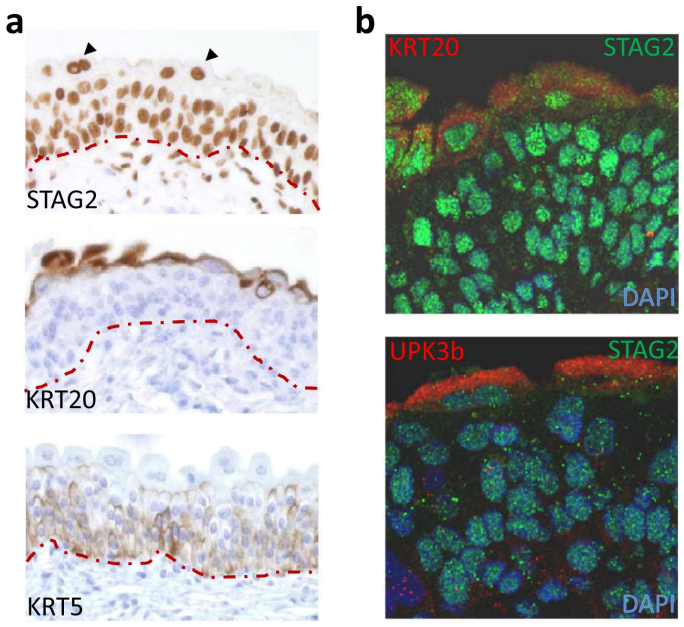

Figure 27. STAG2 is expressed in all layers of normal urothelium. (a) STAG2 protein is detected throughout normal human urothelium, including umbrella cells, which also express KRT20, as well as basal and intermediate cells, positive for KRT5. The arrowheads point to umbrella cells positive for STAG2. The dotted red line marks the location of the basement membrane. (b) STAG2 (green) colocalizes with KRT20 (upper panel, red) and UPK3b (lower panel, red), both markers of umbrella cells.

The cohesin complex can contain either STAG2 or STAG1; therefore, we investigated the relationship between the expression of both proteins in a TMA containing 94 UBCs of all stages and grades. Unsupervised clustering of STAG1 and STAG2 histoscores revealed that most tumors lacking STAG2 expression express STAG1, supporting the existence of a functional cohesin complex in STAG2-negative tumors (Fig. 28). Interestingly, 6 tumors lost both STAG1 and STAG2 expression, all of which were *FGFR3* wild type and of high grade (Fig. 28).

3.5. STAG2 loss of expression is associated with good prognosis

Next, we analyzed the association of lost STAG2 expression with patient outcome using Kaplan-Meier plots, univariable, and multivariable analysis. In addition, we performed the same analyses after stratification. In NMIUBC, patients whose tumors showed loss of STAG2 expression exhibited a significantly lower risk of tumor progression (hazard ratio (HR) 0.27, P=0.007) but not of recurrence (Fig. 29a, b, Tables 15, 16). In the multivariable analysis, adjusting for region, age, stage, grade, multiplicity, number of recurrences and treatment, STAG2 loss was not an independent predictor of outcome in NMIUBC: the HR for progression- free survival was 0.56, P=0.163 (Fig 29b, Table 16). In the case of MIUBC, Kaplan-Meier analyses did not yield a significant association between loss of STAG2 and patient outcome (Fig. 29 c, d, Tables 17, 18). However, multivariable analysis adjusting for region, tumor stage, number of affected nodes, metastasis and treatment showed that loss of STAG2 expression was an independent predictor of cancer-specific survival (HR=0.44, P=0.018) (Fig. 29d, Table 18).

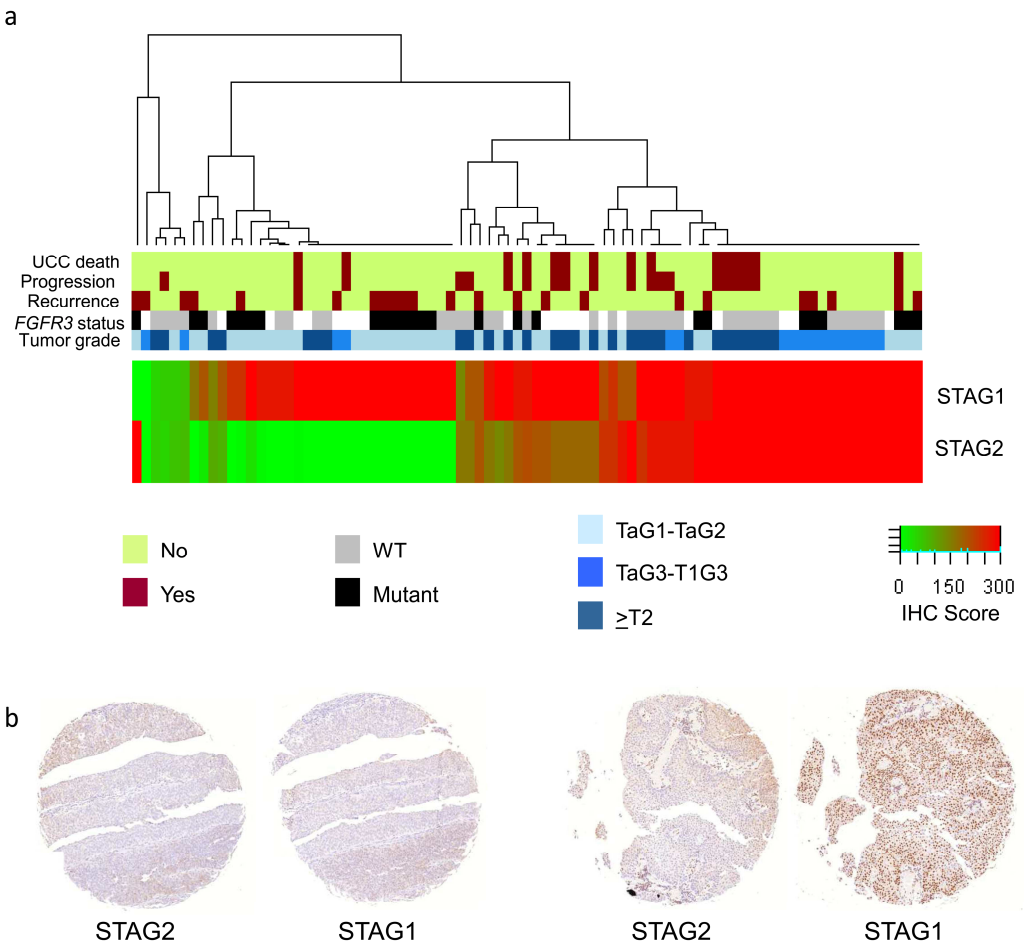

Figure 28. Expression of STAG1 and STAG2 in UBC. (a) Non-hierarchical clustering of STAG1 and STAG2 levels. IHC scores are showed in a green (lowest)-red (highest) color code. Bars below the dendogram include information about prognosis (pistachio/bourbon), tumor stage and grade (tones of blue), and *FGFR3* mutational status (grey/black). White squares indicate information is not available. (b) Examples of one of the few tumors losing expression of both STAG1 and STAG2 (left), and another case displaying loss of STAG2 but positive staining of STAG1 (right).

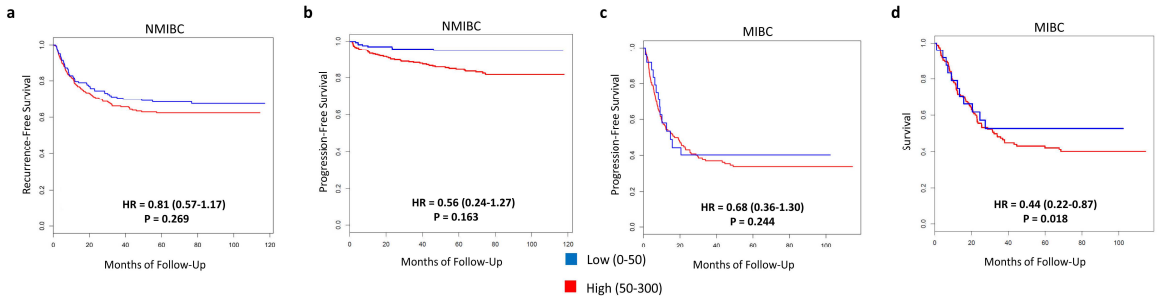

Figure 29. Kaplan-Meier plots showing association between STAG2 expression and outcome. Expression was defined by histoschore (see Materials and methods). (a) Recurrence in NMIBC patients expressing high (n=309) and low (n=171) levels of STAG2. (b) Progression in NMIBC subjects displaying high (n=309) and low (n=171) levels of STAG2. (c) Progression in MIBC individuals presenting high (n=158) and low (n=24) levels of STAG2. (d) UBC-specific survival in MIBC patients expressing high (n=158) and low (n=24) levels of STAG2. P values correspond to the results of multivariable analyses, details on which are given in Tables 22-25.

Table 15. **Association between STAG2 expression and risk of recurrence in NMIUBC**. Kaplan-Meier and Cox regression multivariable analyses.

			Unadjusted model				Adjusted model		
STAG2 histoscore	**Events N(%)**	**Censored N(%)**	**HR**	**95% CI**	**P-value Cox PH**	**P-value Log-Rank**	**HR**	**95% CI**	**P-value Cox PH**
50-300	101(32.7)	208(67.3)	1 (Ref)	0.6 - 1.16	0.283	0.282	0.81	0.57 - 1.17	0.269
0-50	52(30.4)	119(69.6)	0.83						

Models adjusted for region, gender, TG, multiplicity, tumor size and treatment.

Table 16. **Association between STAG2 expression and risk of progression in NMIUBC.** Kaplan-Meier and Cox regression multivariable analyses.

			Unadjusted model				Adjusted model		
STAG2 histoscore	**Events N(%)**	**Censored N(%)**	**HR**	**95% CI**	**P-value Cox PH**	**P-value Log-Rank**	**HR**	**95% CI**	**P-value Cox PH**
50-300	48(15.5)	261(84.5)	1 (Ref)	0.13 - 0.56	**0.001**	**0.0002**	0.56	0.24 - 1.27	0.163
0-50	8(4.7)	163(95.3)	0.27						

Models adjusted for region, age, TG, multiplicity, number of recurrences and treatment.

Table 17. **Association between STAG2 expression and risk of progression in MIUBC.** Kaplan-Meier and Cox regression multivariable analyses.

			Unadjusted model				Adjusted model		
STAG2 histoscore	**Events N(%)**	**Censored N(%)**	**HR**	**95% CI**	**P-value Cox PH**	**P-value Log-Rank**	**HR**	**95% CI**	**P-value Cox PH**
50-300	96(60.8)	62(39.2)	1 (Ref)	0.51 - 1.55	0.671	0.671	0.68	0.36 - 1.3	0.244
0-50	14(58.3)	10(41.7)	0.89						

Models adjusted for region, tumor stage, number of affected nodes and treatment.

Table 18. **Association between STAG2 expression and risk of UBC-related death in MIUBC.** Kaplan-Meier and Cox regression multivariable analyses.

			Unadjusted model				Adjusted model		
STAG2 histoscore	**Events N(%)**	**Censored N(%)**	**HR**	**95% CI**	**P-value Cox PH**	**P-value Log-Rank**	**HR**	**95% CI**	**P-value Cox PH**
50-300	82(51.9)	76(48.1)	1 (Ref)	0.42 - 1.48	0.463	0.462	0.44	0.22 - 0.87	**0.018**
0-50	11(45.8)	13(54.2)	0.79						

Models adjusted for region, tumor stage, number of affected nodes, metastasis and treatment.

3.6. STAG2 loss of expression is not associated with aneuploidy neither in tumors nor in cell lines

The known role of STAG2 in cohesin, together with some preliminary experimental analyses, has prompted the proposal that STAG2 loss promotes aneuploidy and that this mechanism would contribute to tumor development/progression in glioblastoma, melanoma and Ewing sarcoma (Solomon et al. 2011). To determine whether the same holds for UBC, we used BAC arrays or high-resolution SNP arrays to analyze gene/chromosome copy number changes in 23 non-aggressive UBCs. Out of the 11 tumors lacking STAG2 expression, 9 (81.8%) exhibited normal chromosomal content; the remaining 2 tumors had lost one copy of chromosome 9 (Fig. 30, table

Table 19. **STAG2 expression histoscore and aneuploidy in non-aggressive UBC.**

ID*	TG	STAG2 histoscore[$]	Aneuploidy
30104813*	Ta, PUNLMP**	30	0
10091210*	TaG1	1	0
10093410*	TaG1	0	0
10093510*	TaG1	160	0
30107610*	TaG1	60	- chr. 9
30106815*	TaG2	180	- chr. 9
30108611*	TaG2	60	0
30110214*	TaG2	15	0
30106712*	TaG2	0	- chr. 9
10090110*	TaG1	0	0
10092810*	TaG2	0	- chr. 9
51109411	TaG1	240	0
41101319	TaG1	40	0
41101917	TaG1	125	0
23107618	TaG1	270	0
11106713	TaG1	0	0
20110017	TaG1	160	0
23110614	TaG1	180	0
40101419	TaG1	0	0
20100519	TaG1	0	0
50107313	TaG1	135	- chr. 9
51100917	TaG1	270	0
11111319	TaG1	270	0

* Asterisk refers to samples analyzed for genomic changes using the Illumina HumanHap 1M BeadChip SNP arrays; the remaining samples were analyzed using BAC arrays

** PUNLMP, Papillary urothelial neoplasm of low malignant potential

[$] Histoscore range: 0-300

19). Similarly, 9 of 12 (75%) cases expressing STAG2 were diploid; the remaining 3 cases had also lost chromosome 9. These results indicate that, in UBC, loss of STAG2 does not result in

aneuploidy, in agreement with our finding that loss of STAG2 expression is more frequent in tumors of low grade and stage that are known to be genomically stable (Blaveri et al. 2005; Knowles 2008; Lindgren et al. 2010).

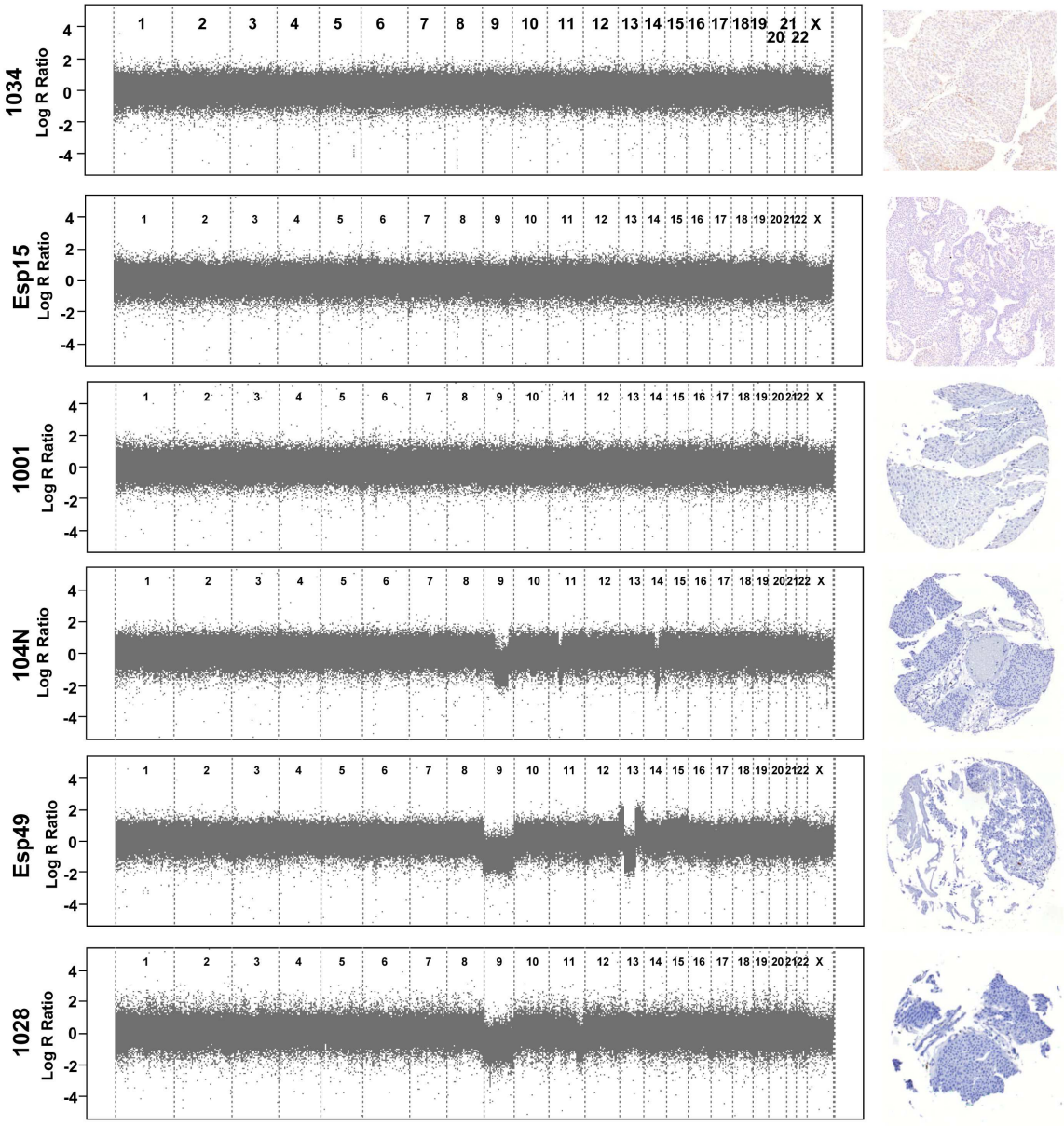

Figure 30. Loss of STAG2 expression is not associated with aneuploidy in primary UBC. Genomic profiles (Illumina HumanHap 1M BeadChip arrays) of low-risk, non-aggressive NMIUBC lacking STAG2 expression. The top 3 tumors (1034, Esp15, and 1001) do not exhibit gene copy number changes; tumor 104N shows interstitial losses at chromosomes 9, 11, and 14; case Esp49 exhibits loss of chromosome 9 and interstitial gains and losses of chromosome 13; and tumor 1028 displays loss of chromosome 9 and interstitial losses at chromosome 11. Corresponding STAG2-negative IHC stainings are shown on the right.

To determine whether STAG2 loss in cultured bladder cancer cells increases aneuploidy, we used lentiviral shRNA constructs to silence STAG2 expression. Twenty-seven out of 33 (81.8%) UBC cell lines express STAG2 at the protein level (Fig. 19a, Fig. 31a). We knocked down STAG2 in 3

cell lines displaying a broad range of phenotypes: UM-UC-11 grow slowly and are wild type for all of the genes involved in UBC tested, RT112 cells harbor an *FGFR3* amplification, and 639V cells grow rapidly, exhibit genomic instability, and also harbor mutations in *PIK3CA, PTEN*, and *TP53* (Julie Earl, unpublished). Despite the fact that the knockdown efficiency was good, we did not find consistent effects on chromosome numbers of metaphase-arrested cells (Fig. 31b-d, Table 20). We performed 3 independent experiments and counted chromosome number in ≥50 metaphase-arrested nuclei for each shRNA condition. Changes in mean and median chromosome numbers upon STAG2 silencing were very small (≤1/metaphase) and inconsistent both when considering the 2 independent shRNAs that reduced STAG2 expression to similar levels and among different

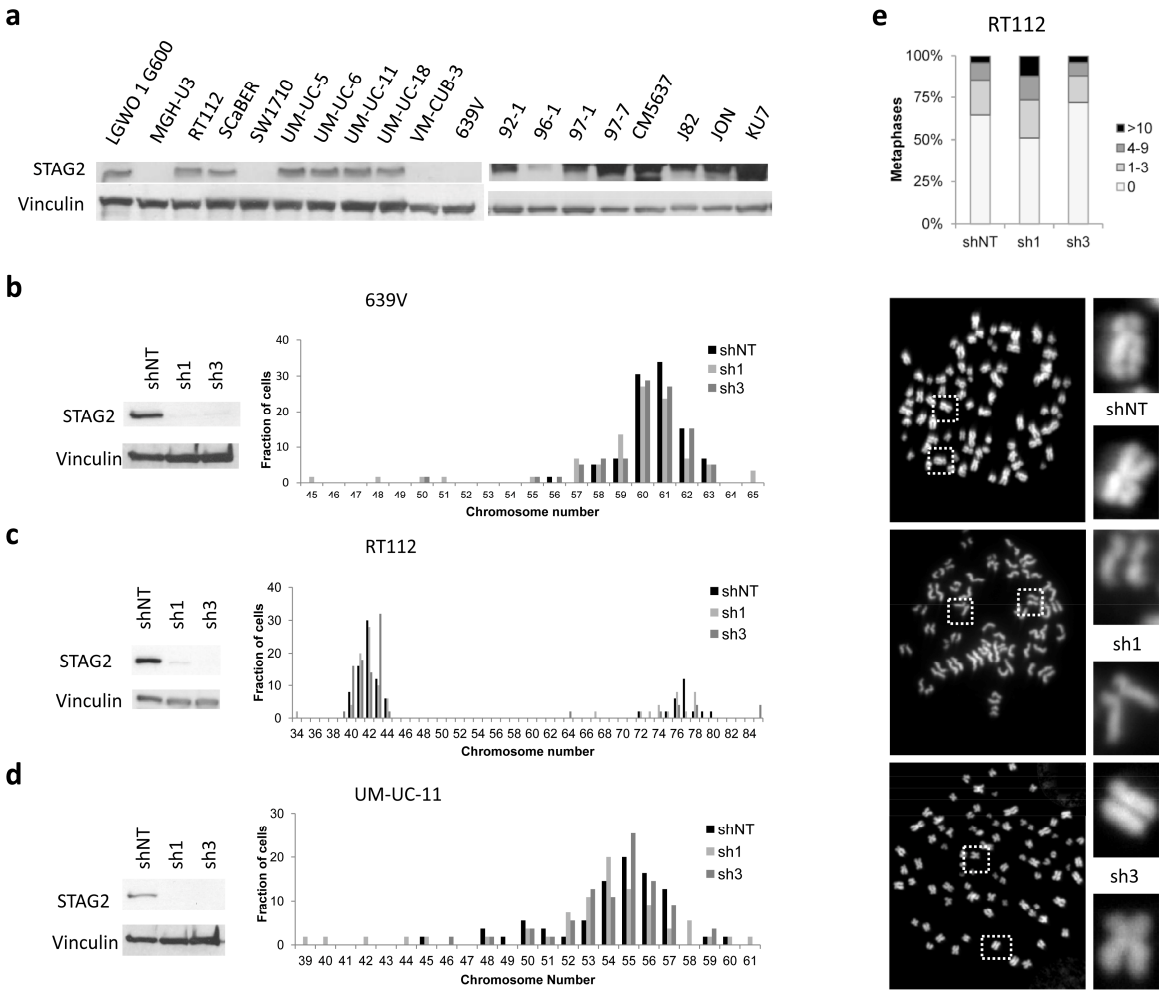

Figure 31. Loss of STAG2 is not associated with aneuploidy in vitro. (a) Immunoblotting analysis of STAG2 in UBC cell lines shows undetectable expression in 4 of the cell lines used for functional studies. (b-d) Efficient STAG2 knockdown was achieved in three UBC cell lines, as demonstrated by immunobloting. This did not lead to consistent changes in chromosome number (quantification shown in Table 27). (e) Depletion of STAG2 in RT112 cells did not cause consistent changes in the occurrence of aberrant centromeric cohesion. The grey color code refers to the number of chromosomes with defective centromeric cohesion in a given metaphase. shNT: non-targeting short hairpin RNA.

cell lines. The experimental differences observed were not statistically significant with the exception of sh1 in 639V (which exhibited an average of 60.61 chromosomes per metaphase in the control-shRNA treated cells, vs. 59.39 in cells infected with sh1, P= 0.022) (Table 20).

Since STAG2-cohesin has been shown to be responsible for centromeric cohesion (Canudas and Smith 2009; Remeseiro et al. 2012a), we also measured centromeric cohesion defects upon STAG2 knockdown in RT112 cells, but found no significant changes in cohesion (Fig. 31e).

3.7. STAG2 overexpression and silencing in vitro impairs cellular growth

To explore additional mechanisms through which STAG2 could contribute to tumorigenesis, we assessed proliferation and colony formation upon STAG2 knockdown *in vitro*. Surprisingly, a consistent and significant decrease in colony formation was observed in 5 different UBC cell lines (Fig. 32a-e). The strongest effect was observed in RT112 cells, where knockdown resulted in a 81 and 82% reduction in colony number with two different shRNAs (P=0.006 and 0.005), respectively (Fig. 32a). The mildest effect was detected in UM-UC-11 cells, where the reduction was of 28% with both shRNAs, although the difference was statistically significant only for one of them (P=0.02 and 0.06) (Fig. 32e). This is at odds with the extensive evidence implicating STAG2 function as a tumor suppressor in UBC. Therefore, we introduced wild type *STAG2* cDNA in 3 cell lines lacking STAG2: UM-UC-6 harbor R305* (exon 11) and F1228L (exon 33) alterations, VM-CUB-3 harbor a 10-bp deletion in exon 6, and LGW0 1 G600 cells are wild-type in exons 3–35. In all three cases, STAG2 rescue also resulted in decreased colony formation (Fig. 32 f-h).

Table 20. **Relationship between STAG2 knock-down and chromosome number in metaphase-arrested bladder cancer cell lines.**

Cell line	n		shNT	sh1	sh3
		Mean chromosome number	60.61	59.39	60.05
639v	59	Median	61	60	60
		P (Wilcoxon)	-	**0.022**	0.24
		Mean chromosome number	51.64	51.74	47.4
RT112	50	Median	42	42	42.5
		P (Wilcoxon)	-	0.86	0.25
		Mean chromosome number	54.24	53.24	54.34
UM-UC-11	50	Median	55	54	55
		P (Wilcoxon)	-	0.2	0.81

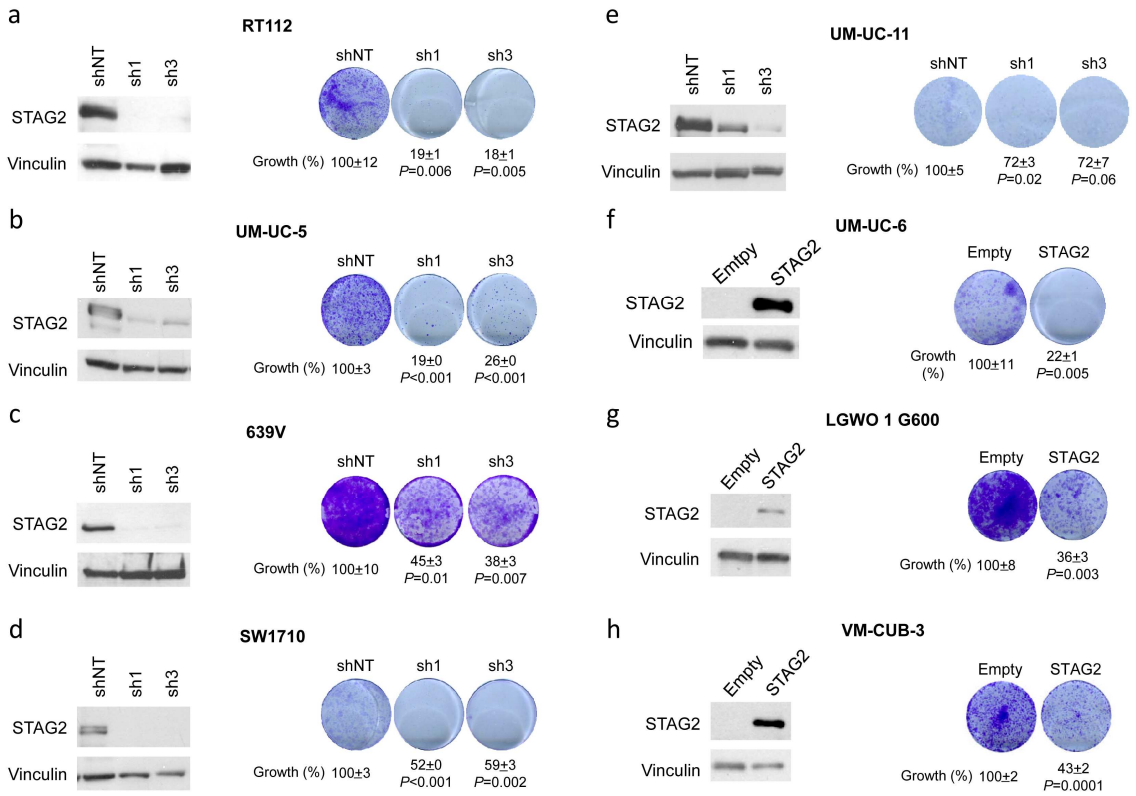

Figure 32. *In vitro* knockdown and re-expression of STAG2 leads to impaired colony formation ability. (a-e) Efficient STAG2 knockdown, as demonstrated by immunoblotting, consistently and significantly associates with reduced colony formation in RT112, UM-UC-5, 639V, SW1710, and UM-UC-11 cell lines. (f-h) Efficient STAG2 re-expression, as demonstrated by immunoblotting, consistently and significantly associates with reduced colony formation in UM-UC-6, LGWO 1 G600, and VM-CUB-3 cell lines. Growth was quantified in triplicate experiments and is expressed in percentage ± Standard Error of the Mean. shNT: non-targeting short hairpin RNA; Empty: empty overexpression vector; STAG2: STAG2-expressing vector.

3.8. Knocking down STAG2 in UBC cell lines does not significantly affect transcriptional profiles

The lack of significant effects on chromosome cohesion and ploidy prompted the possibility that STAG2 might contribute to UBC development through its role as a regulator of chromatin and gene expression. Therefore, we effectively knocked down STAG2 in 2 cell lines (RT112 and 639V) using lentivirus coding for control shRNA or 2 different shRNA constructs targeting STAG2 (Fig. 33a), and analyzed the transcriptome using RNA-Seq (Table 21). Principal component analysis disclosed a clear separation between the two cell lines but, within each line, samples infected with control and the two STAG2-targeting shRNAs were very close to each other (Fig. 33b). In agreement with this observation, only 0.2% - 0.8% genes were significantly regulated (Table 21). We considered a gene to be differentially expressed when Q<0.05 (where Q-values correspond to the false discovery rate-adjusted P values).

There were no genes regulated that were common between the 2 shRNAs and the 2 cell lines. A less stringent analysis, focused on genes that showed an expression fold-change equal or greater to that of STAG2 in the different conditions compared to the control shRNA (Table 22), yielded 126 genes altered in both cell lines with both shRNAs (Fig. 33c). We selected 25 genes for RT-qPCR validation, including STAG2: 6 genes that were regulated in both cell lines, 8 changing

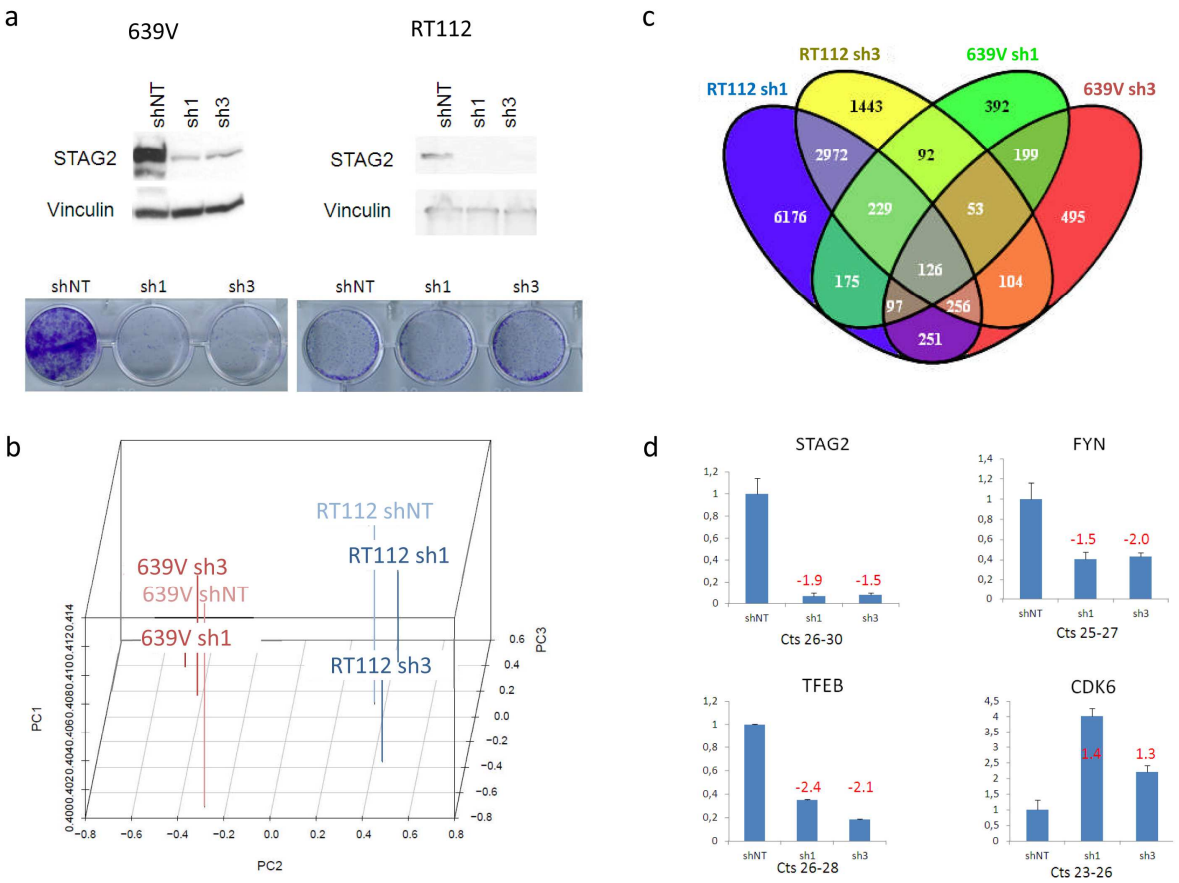

Figure 33. RNA-sequencing upon STAG2 silencing in UBC cell lines. (a) Efficient STAG2 silencing leading to decreased colony formation capabilities was achieved in 639V and RT112. (b) 3D Principal Component Analysis shows distinct separation between both cell lines; within each cell line, there is a slight separation between control shRNA (shNT) and the shRNas targeting STAG2 (sh1 and sh3). PC: principal component. (c) Venn diagram showing the number of common genes between the RT112 and 639V cell lines upon STAG2 silencing; each data sets includes the genes up or downregulated at fold-changes ≥STAG2 comparing a given silencing construct (sh1 or sh3) with the control (shNT) for each cell line. 126 genes were found to be altered in both cell lines using both shRNA constructs. (d) A subset of genes was chosen for RT-qPCR validation; some were common for both cell lines and some were found only in one. Genes up and downregulated to different extents were chosen to ensure the validation subset was as representative as possible. Shown are RT-qPCR results for four genes validated in independent biological triplicates for 639V. The fold-change value obtained in the RNA-sequencing study is shown in red.

only in 639V, and 10 that exhibited different expression solely in RT112[1]. We also made sure that the genes belonged to different pathways and ensured the inclusion of genes that were both up and down-regulated, to get a representative overview of the sequencing quality. The regulation of 22/25 of the genes assayed was confirmed using RNA from the same experiment used for RNA-Seq. However, when mRNA from independent experiments was used, only 8 genes were consistently regulated. STAG2 was the only one downregulated in both cell lines. In 639V, FYN and TFEB were downregulated and CDK6, TSC22D2, and CD83 were upregulated; in RT112, WNT7A was downregulated and KRT14 was upregulated (Fig. 33d).

Table 21. **Summary of the RNA sequencing study.** RIN: RNA Integrity Number. % Align (PF): The percentage of filtered [Passing Filter] reads that were uniquely aligned to the reference. % >=Q30 bases (PF): Yield of bases with Q30 or higher from clusters passing filter divided by total yield of clusters passing filter. Illumina Q30 quality score is associated with an error rate of 0.001 and therefore a 99.9% inferred accuracy for base calling.

Run type	Read length	Sample	RIN	Number of high quality reads (Clusters PF)	Yield (Mbases)	% of reads mapped confidently (% Align PF)	% >=Q30 bases (PF)
Single Read	40	639v-p27-shNT	6.8	17,885,709	715	72.6	91.4
		639v-p27-STAG2-sh1	8.5	18,104,376	724	70.6	91.3
		639v-p27-STAG2-sh3	7.8	17,691,116	708	74.1	91.3
		RT112-p33-shNT	9.1	17,749,443	710	75.8	93.4
		RT112-p33-STAG2-sh1	6.8	18,372,849	735	71.6	93.4
		RT112-p33-STAG2-sh3	7.7	16,390,860	656	73.2	93.4

Table 22. **Results of the RNA sequencing analysis.**

		shNT vs sh1	**shNT vs sh3**
639v	# significant differentially expressed genes	75/17069	109/17069
	%	0.4	0.6
	fold-change range of DEGs (abs)	1.3 - 7.7	1.4 - 7.1
	STAG2 log2 fold-change	-1.9	-1.5
RT112	# significant differentially expressed genes	148/17678	27/17787
	%	0.8	0.2
	fold-change range of DEGs (abs)	1.5 - 7.9	1.7 - 11.2
	STAG2 log2 fold-change	-0.6	-0.8

[1] We tested FYN, TSC22D3, NPR2, TFEB, CDK6, MFGE8, CD83, and MAPK15, which were regulated only in 639V, PIM1, MET, KRT14, IGF1R, WNT6, WNT7A, WNT9A, and GLI2 in RT112, and STAG2, TSC22D2, RHOBTB3, MTMR3, IFI6, CCDC88B, and COL1A1 in both.

Discussion

Despite its high incidence and cost, the molecular landscape of UBC remains relatively understudied. We used WES to identify novel genes involved in UBC. We were aware of a major sequencing initiative of MIUBC by the CGARN, so we focused our analyses on NMIUBC.

1. THE MUTATIONAL LANDSCAPE OF UBC

Five other UBC sequencing initiatives have been published besides our study (Gui et al. 2011; Guo et al. 2013; CGARN 2014a; Cazier et al. 2014; Nordentoft et al. 2014) (Table 23), the first 3 of them focusing on MIUBC. As expected, the WGS study identified more SNVs, even when taking into account only exonic regions (Cazier et al. 2014). Gui et al. reported a much lower mutation number than the rest of studies, but this is accounted for by the fact that they reported somatic mutations after filtering SNPs out (2011). However, the CGARN study identified 1.8x more mutations per patient as we did (2014a). The differences in mutation number in our study could be accounted by the fact that we included NMI tumors, as well as by the relatively low number of cases of NMIUBC studied. However, other reasons likely contribute to discrepancies between the studies by Nordentoft et al., Guo et al. and the CGARN. All studies used Illumina HiSeq for sequencing, but data were processed differently. The CGARN SNV calling strategy did not include filtering out known germline variants, possibly accounting for the observed differences, especially when we take into account that the CGARN used MutSig (Lawrence et al. 2013) to call for mutations, taking into account background mutation rate, a process that should yield a lower number of false-positive SNVs called than traditional strategies.

Table 23. **Summary of NGS studies of UBC.** WES: Whole Exome Sequencing. WGS: Whole Genome Sequencing. *Known SNPs were filtered out.

Study	Study type	Total cases (NMI-LR, NMI-HR, MI)	NMI LR cases	NMI HR cases	MI cases	Average somatic mutations per patient (exonic)
This study	WES	17	5	10	2	169
Gui et al. 2011	WES	9	0	0	9	52*
Guo et al. 2013	WES	99	5	33	61	114
CGARN 2014q	WES	130	0	0	130	302
Cazier et al. 2014	WGS	14	4	5	5	417
Nordentoft et al. 2014	WES	38	20		18	195

A recent meta-analysis of 27 tumor types sequenced mainly at the Broad Institute revealed a >1,000 fold variation in the number of non-synonymous variants per tumor depending on the tumor type (or site) (Lawrence et al. 2013). Pediatric tumors exhibited the lowest number of mutations whereas cancers known to be associated with environmental carcinogens, such as

those from the lung or melanoma, presented the highest number. The frequencies identified by Guo, Nordentoft, and our group fell in the mid-high range among solid tumors of the adult; the CGARN frequencies approached the rates found in lung and melanoma tumors.

Consistently, the 6 bladder cancer studies identified C>T/G>A transitions as the most common nucleotide change in exons. C mutations in a TpC dinucleotide context have been recently shown to occur frequently in several cancer types (Lawrence et al. 2013) and have been proposed to be caused by APOBEC, enzymes that deaminate cytosine nucleotides in single stranded DNA, giving rise to uracil that is excised by uracil DNA glycosylases. This process yields an apurinic site, opposite to which an A can be inserted, resulting in C>T transitions (Loeb and Preston 1986). The comprehensive study of 38 UBCs by Nordentoft and colleagues showed that around 1/3 of UBCs exhibit an APOBEC mutational pattern (2014). Moreover, *APOBEC3B* levels have been shown to be upregulated in bladder tumors exhibiting the APOBEC mutational signature (Burns et al. 2013; Roberts et al. 2013).

All the studies included a small number of tumors with very few mutations, mainly in genes different from those reported to be frequently altered in UBC or newly identified in these studies. Technical limitations may contribute to this observation, including those derived from sequencing only the coding part of the genome. Other kinds of driver mutations, such as those mapping to non-coding regions, or alterations involving larger chromosomal stretches (chromosomal rearrangements or copy number changes), might play a driver role in those cases. Alternatively, it is possible that a subgroup of UBC develops mainly through epigenetic alterations that have not been analyzed so far, such as is the case of medulloblastoma (Hovestadt et al. 2014).

2. *ARID1A* MUTATIONS AND EXPRESSION IN UBC

NGS studies have uncovered a previously unsuspected role of genes coding for histone modifiers and chromatin remodelers acting as tumor suppressors in many human cancers (reviewed in Wilson and Roberts 2011). A recent meta-analysis showed that the SWI/SNF is the most commonly altered chromatin-regulatory complex in human tumors (Kadoch et al. 2013). ARID1A was identified as the most frequently mutated SWI/SNF subunit. *ARID1A* somatic mutations were first identified in ovarian carcinomas (Jones et al. 2010; Wiegand et al. 2010). Shortly after, mutations were reported in uterine endometrioid tumors, esophageal and gastric carcinomas, hepatocellular tumors, and other less common malignancies, at frequencies of 10-40% (reviewed in Wu 2014). In 2011, mutations in *ARID1A* and other chromatin remodelers were reported in the first UBC WES study of mostly aggressive tumors (Gui et al. 2011). Besides being

inactivated by mutations, ARID1A was also shown to be inactivated by epigenetic silencing in invasive ductal breast tumors (Zhang et al. 2013).

In this study, we found that 11.2% of UBC harbor *ARID1A* pathogenic mutations. This mutation rate is lower than the overall 21.4% rate identified in 7 independent studies of this tumor (Gui et al. 2011; Guo et al. 2013; Cazier et al. 2014; CGARN 2014a; Nordentoft et al. 2014; Ross et al. 2014; Kim et al. 2014). However, our analysis included 36% of non-aggressive tumors, as opposed to only 6.7% of the cases analyzed by the other groups (Table 8). In our study, all of the truncating mutations identified were found in aggressive tumors. Joint analysis of the eight studies shows that *ARID1A* relevant mutations occur in 14.4% of NMI and 23.6% of MI cases (P=0.02) (Supp. Table 12). Overall, 54/66 (81.8%) mutations found in aggressive tumors were truncating, compared to 5/9 (55.6%) in non-aggressive tumors (P=0.09). In agreement with this observation, we found ARID1A protein expression to be lost significantly more frequently in aggressive tumors. Interestingly, we identified 1 patient and 1 cell line harboring multiple relevant mutations; Guo et al. also identified 2 patients with multiple mutations (2013). This could be indicative of biallelic inactivation of the gene, as well as of intratumoral heterogeneity, both common events in other tumor types (Jones et al. 2010; Guan et al. 2011a).

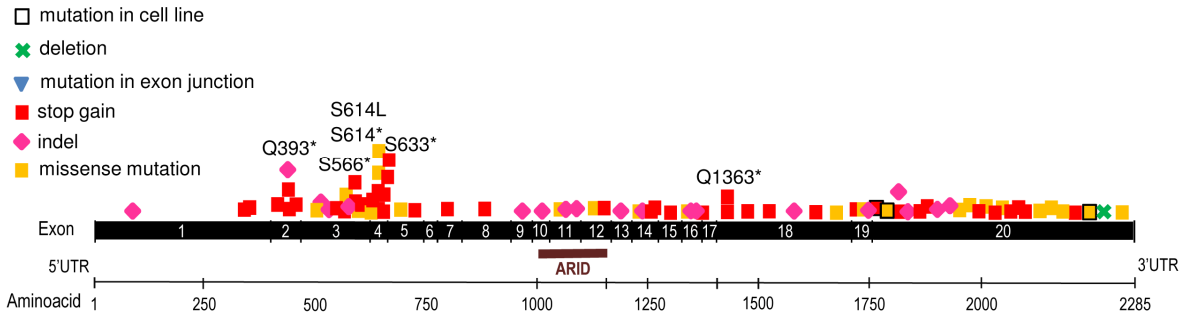

Figure 34. Summary of type and distribution of *ARID1A* nonsynonymous somatic mutations found in all published studies. The amino acid changes in mutations occurring more than once are specified above their location.

Taking into account all studies for which information was available, we did not identify significant differences in the frequency of *ARID1A* mutations according to sex, age, or smoking status, although there was a tendency for *ARID1A* mutations to be more common in females than males in the five studies for which this information was available (23.8% vs. 16.9%) (P=0.2) (Supp. Table 12). Interestingly, there is evidence that women present with more aggressive tumors and that they have a worse outcome than men when adjusting for stage and grade (Cook et al. 2011). A summary of all reported identified mutations can be found in Figure 34. Although mutations were distributed all along the protein, there was a cluster of truncating mutations in exons 3-4, although no known functional domain maps to that location.

3. *ARID1A* ALTERATIONS IN THE CONTEXT OF UBC'S GENETIC LANDSCAPE

We also explored the relationship between mutations in *ARID1A* and in other genes involved in UBC. Mutual exclusivity analyses of all studies published to date suggested that alterations in *ARID1A* were exclusive with alterations in *FGFR3* (OR=0.4, P=0.007), and *STAG2* (OR=0.4, P=0.03) (Supp. Table 13). In agreement with this, we found that *ARID1A* mutations usually occurred in poor-prognosis *FGFR3* wild type tumors. Consistently, we have shown that ARID1A and FGFR3 immunoscores exhibited a significant direct correlation.

We did not find any relationship between ARID1A expression and nuclear p53 staining (associated with *TP53* alterations). Similarly, in a joint analysis of all available studies we found no association between *TP53* and *ARID1A* alterations (OR=1.3) (Supp. Table 13), suggesting that *ARID1A* inactivation is an alternative pathway of UBC progression. Similarly, no association between *ARID1A* and *TP53* alterations has been detected in other tumor types (Wang et al. 2011; Bosse et al. 2013; Kadoch et al. 2013; Allo et al. 2014), suggesting a broader functional significance of this observations. *In vitro* studies in ovarian cancer cells have shown that ARID1A and p53 co-immunoprecipitate and are both required for p21-mediated cell cycle arrest (Guan et al. 2011b), suggesting that the tumor suppressor functions of ARID1A and p53 converge in multiple human tumors. Accordingly, we found ARID1A expression to be associated with worse patient outcome. Further studies are required to ascertain the functional relationship of these two tumor suppressors.

Regarding the sequence of mutational events, Nordentoft et al. have proposed - on the basis of sequencing studies of 4 pairs of primary NMIUBC-progressor tumors - that mutations in chromatin modifiers (*KDM6A, MLL2, ARID1A*) occur early in tumor development and associate with worse prognosis (2014). Moreover, several pretumoral lesions have been reported to exhibit *ARID1A* mutations in endometriotic cysts, endometriosis, endometrial hyperplasia, and normal squamous esophageal epithelium (Ayhan et al. 2012; Yamamoto et al. 2012; Werner et al. 2013; Streppel et al. 2014). Although most of the precursor lesions analyzed are related to gynecologic cancers, this suggests that *ARID1A* alterations causing SWI/SNF inactivation might be an early event in carcinogenesis, at least in some instances.

Co-occurrence analyses of all published reports suggested that alterations in *ARID1A* associated with alterations in *KRAS*, which was mutated in a subset of non-aggressive UBC (OR=5.2, P=0.02) (Supp. Table 13). Moreover, us an others have identified a small subset of non-

aggressive UBCs with mutations in *ARID1A*. Altogether, we propose that *ARID1A* alterations that occur early during UBC evolution are important in a subset of poor-prognosis *TP53* wild type tumors belonging to the aggressive pathway of UBC development. Additionally, KRAS-mutant, non-aggressive tumors might progress to MI status via the acquisition of further alterations in ARID1A.

4. ARID1A: A ROLE IN THE CONTROL OF DIFFERENTIATION AND PROLIFERATION?

Despite the wealth of data on mutations in SWI/SNF chromatin remodelers, the mechanisms through which these act as tumor suppressors in human cancer remain unclear (Wilson and Roberts 2011). This is especially true for ARID1A, whose function within the BAF SWI/SNF complex is not well characterized.

The SWI/SNF complex participates in the control of tissue-specific gene expression programs by regulating promoter accessibility to transcription factors and epigenetic modifiers. *ARID1A* mutations might contribute to global epigenetic changes upon BAF inactivation. The region of ARID1A that interacts with DNA has been shown to be required for BAF complexes to bind to nucleosomes in MEFs (Chandler et al. 2013). Therefore, BAF-mediated nucleosomal remodeling could be globally altered upon *ARID1A* mutation, rendering CpG islands and gene promoters accessible to the action of epigenetic modifiers. Supporting this, *ARID1A* mutations have been found in colorectal tumors with abnormal CpG island hypermethylation (Tahara et al. 2014). Moreover, Duymich and colleagues have reported preliminary results suggesting that UBC cells harboring *ARID1A* mutations exhibited differentially methylated CpG sites (2013). Given the mutual exclusivity between *ARID1A* and *FGFR3* mutations we have identified, it is tempting to hypothesize that alterations in *ARID1A* might be involved in the multiple regional epigenetic silencing phenotype described in *FGFR3*-wild type, aggressive UBC (Vallot et al. 2011).

This effect could be particularly important in the context of cell differentiation and proliferation. The BAF subtype of SWI/SNF complexes can contain either BRG1 or BRM, and either ARID1A or ARID1B. SWI/SNF subunits have been shown to facilitate MEF reprogramming: cell fractionation identified several BAF components significantly enriched in reprogramming-competent extracts, including Brg1, Baf155, Arid1a, and Arid1b (Singhal et al. 2010). Specifically, Brg1 and Baf155 were shown to mediate faster demethylation of Oct4 and Nanog promoters, thus enhancing reprogramming (Singhal et al. 2010).

BRG1 is the best studied BAF component with regards to differentiation. The group of Sánchez-Céspedes has focused on studying BRG1 alterations in the context of lung cancer. They first identified common BRG1 homozygous mutations in lung cancer cell lines (Medina et al. 2008) and proceeded to confirm those findings in lung tumors (Rodríguez-Nieto et al. 2011). They then demonstrated that re-introducing wild type, but not mutant BRG1, in lung cancer cells lacking the protein switched gene expression to a signature typical of normal lung (Romero et al. 2012). Interestingly, among the differentially expressed genes they found MYC targets. They proceeded to demonstrate that BRG1 binding to regions near the promoter of MYC targets repressed their expression.

This repression could be accounted for by global epigenetic changes upon BRG1 inactivation: the CpG islands of MYC targets could become methylated in the absence of functional BRG1-BAF. Myc is unable to bind its targets when its DNA binding sequences (CACGTG, called E boxes) are methylated (Prendergast et al. et al. 1991; Perini et al. 2005). Therefore, inactivating mutations in *BRG1* causing E box methylation could result in repression of the MYC transcriptional programme.

It is tempting to speculate that alterations in other BAF components, such as ARID1A, also result in the maintenance of undifferentiated gene expression, possibly also via the control of MYC activity. In fact, Arid1a has been shown to directly repress c-Myc expression, contributing to cell cycle arrest-dependent differentiation in an osteoblastic mouse cells (Nagl et al. 2006). In humans, loss of ARID1A expression is more common in poorly differentiated gastric tumors than in well-differentiated ones (Wang et al. 2012). Interestingly, mutations in *ARID1A* and *BRG1* are mutually exclusive in high grade UBCs (Kim et al. 2014), suggesting that disruption of either of them affects the function of the SWI/SNF complex. In our IHC analyses, the cluster of ARID1A-negative aggressive tumors also exhibited low levels of KRT20, and a subset of these tumors expressed mid to high levels of KRT5/6 and KRT14, suggesting a blockade in urothelial differentiation.

With all this in mind, it seems plausible for *ARID1A* inactivation in UBC to have a tumor suppressive effect that affects global epigenetic status, as well as cell proliferation and differentiation via the control of MYC activity, as shown for BRG1. However, BRG1 is the catalytic subunit of the BAF complex, unlike ARID1A, which does not have catalytic activity but is required for ARID1A-BAF binding to DNA and is believed to influence selectivity of target recruiting. It has been shown that ARID1A and ARID1B have differential activities. In the study mentioned above, knocking down ARID1B did not hinder osteoblast arrest (Nagl et al. 2006). This could imply that, in

the absence of ARID1A, ARID1B-BAF complexes, unable to restrict proliferation and induce differentiation, might become prevalent. Moreover, chromatin immunoprecipitation (ChIP)-sequencing experiments have shown that SWI/SNF is present at promoters of its own subunits and thus can control their relative abundances (Euskirchen et al. 2011). ARID1A inactivation could lead to feedback loops that alter the stoichiometry of BAF subunits, promoting the formation of ARID1B-BAF complexes in lieu of ARID1A-BAF.

ARID1A has also been shown to restrict proliferation in several tumor cells *in vitro* (Wu et al. 2014). Surprisingly, knocking down ARID1A in three different UBC cells resulted in reduced colony formation and very low cell viability. Modest knockdown in gastric (Zang et al. 2012), hepatocellular (Huang et al. 2012), breast (Mamo et al. 2012), bile duct (Chan-on et al. 2013) and ovarian (Guan et al. 2011b) cancer cell lines has been reported to lead to increased proliferation. However, more efficient depletion of ARID1A in immortalized pancreatic and keratinocytic epithelial cells also reduced proliferation (Shain et al. 2012). These differences could be explained by context-specific or dose-dependent effects of ARID1A. In fact, inactivation of a single Arid1a allele is lethal for mouse pre-implantation embryos (Gao et al. 2008), indicating that changes in ARID1A protein dose can be deleterious. Experiments with inducible shRNAs targeting ARID1A, or inducible overexpression, would shed some light on the role of ARID1A in normal urothelial cells. The creation of a urothelial-specific inducible ARID1A KO mouse model would also contribute to test these hypotheses, using a combination of ChIP, bisulfite assays, and expression analyses, either at concrete loci or in a genome-wide manner.

5. *STAG2*: A NEW, COMMON, TUMOR SUPPRESSOR GENE WITH A MAJOR ROLE IN UBC

In 2011, inactivating STAG2 mutations were reported in glioblastoma, Ewing sarcoma, and melanoma, with up to 20% of these tumors exhibiting loss of protein expression using IHC (Solomon et al. 2011). In the last year, our group and others have shown that UBC is the tumor type most commonly harboring *STAG2* alterations (Guo et al. 2013; Solomon et al. 2013; Cazier et al. 2014; CGARN 2014a; Kim et al. 2014; Nordentoft et al. 2014; Taylor et al. 2014).

In this work, we found that 14.3% of all UBCs harbor pathogenic mutations in *STAG2*. This mutation rate fell within the 8.2-25.5% rates identified in 7 independent studies published with ours or subsequently (Table 10) (Guo et al. 2013; Solomon et al. 2013; Cazier et al. 2014; CGARN 2014a; Kim et al. 2014; Nordentoft et al. 2014; Taylor et al. 2014). In our cohort, 22.2% of non-aggressive vs. 10.2% of aggressive tumors were found to be mutated. Joint analysis with the six studies for which stage and grade information was available showed that *STAG2* relevant

mutations were present in 29.6% of non-aggressive and 15.7% of aggressive cases (P<0.001), supporting our findings (Supp. Table 14). In agreement with this, we also found that the frequency of STAG2 protein losses directly associated with the benign nature of the various UBC subtypes.

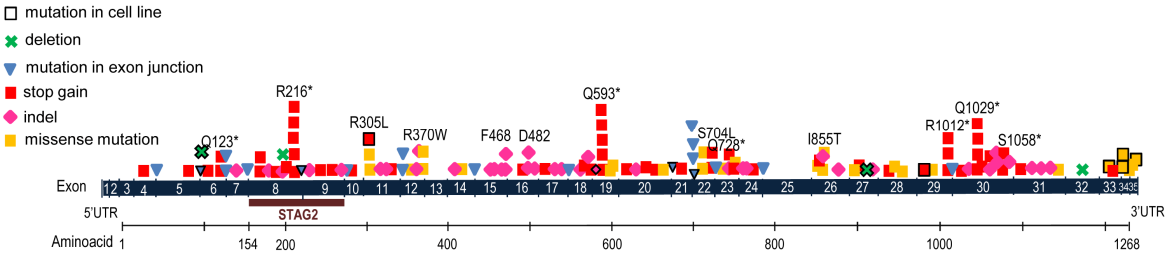

Figure 35. Summary of type and distribution of *STAG2* nonsynonymous somatic mutations found in all UBC published studies. Recurrent mutations are indicated. The amino acid changes in mutations occurring more than once are specified above their location.

Joint analysis of all studies with available information unveiled a tendency for a higher frequency of mutations in tumors from patients <60, females, and in individuals with a history of smoking, although these differences were only significant for gender (P=0.01) (Supp. Table 14). This finding is of interest because *STAG2* is on the X chromosome. Truncating *STAG2* mutations were most common although missense mutations of undetermined biological significance have also been reported (Fig. 35). Overall, all (39/39) truncating mutations (nonsense, frameshifts, and splice-site) led to loss of protein expression, as opposed to none of the missense (0/9) mutations (Solomon et al. 2013; Taylor et al. 2014). Interestingly, missense mutations were more common in aggressive (15/66, 22.7%) than in non aggressive (5/69, 7.2%) tumors (P= 0.02).

The exact timing at which STAG2 mutations occur is not well understood yet. Using IHC, the majority of the tumors analyzed showed homogeneous STAG2 expression. This, together with the high prevalence of mutations in non-aggressive tumors, suggests that STAG2 inactivation is an early ancestral genetic event. We, and others, have reported a small subset of tumors with heterogeneous STAG2 expression (Solomon et al. 2013; Taylor et al. 2014) supporting a role in clonal evolution in some tumors. Moreover, 1/4 matched cases studied by Nordentoft et al. exhibited an indel mutation in STAG2 in the tumor ancestral branch (2014). A small subset of cases harbored more than one STAG2 mutation, possibly suggesting convergent evolution as reported for other tumor suppressors in renal cancer (Gerlinger et al. 2012; Real et al. 2014).

In addition to mutations, *STAG2* inactivation can take place through other mechanisms including genomic deletions, identified by us and by Guo et al. (2013), and epigenetic silencing. Guo and colleagues have reported STAG2 promoter hypermethylation in 7/30 (23%) of tumors (2013). This could account for the 9% of tumors where STAG2 protein was not expressed but we did not find gene mutations. The participation of multiple mechanisms converging in frequent

STAG2 inactivation, and their occurrence early in tumor development, support that *STAG2* is a driver gene in UBC. Studying these in a greater number of cases and relating their prevalence to tumor stage and grade would help to test this hypothesis.

6. PLACING *STAG2* ALTERATIONS IN THE GENETIC LANDSCAPE OF UBC

We have found that STAG2 loss of expression is associated with non-aggressive tumors, which present the highest frequency of *FGFR3* mutations. The analysis of all available data from UBC also showed that mutations in *STAG2* and *FGFR3* frequently co-occur (OR=4.6, P<0.0001) (Supp. Table 15). Mutations in *TP53* and p53 overexpression are frequent in MIUBC. In our analysis, STAG2 loss was associated with normal p53 expression in LR-NMIUBC (OR=4.2, P=0.01) (Table 12). Similarly, mutations in *STAG2* were found to be mutually exclusive with mutations in *TP53* in our analysis of the data by Guo and colleagues, which included mainly aggressive cases (P=0.05, OR<0.0001), and truncating mutations in *STAG2* were associated to wild type *TP53* in another study including cases of all stages and grades (Taylor et al. 2014). However, Solomon and colleagues reported that UBCs frequently exhibit *STAG2* mutations and p53 overexpression in tumors of all stages and grades (2013). A more thorough analysis of these data revealed that 5/10 tumors with STAG2 mutations overexpressing p53 belonged to the aggressive category, which might have acquired *TP53* alterations during tumor progression. More work is needed to place *STAG2* mutations in the context of p53 alterations but we propose that *STAG2* mutations occur mainly leading to non-aggressive tumors, some of which acquire additional alterations in important tumor suppressors - such as *TP53* or *PTEN* - and progress to become muscle-invasive.

At another level, it has been proposed that UBC harbor a mutational signature associated with APOBEC3B cytidine deaminase activity, enriched in C>T transitions (Burns et al. 2013). Nordentoft and colleagues have reported a negative correlation between *STAG2* mRNA levels and the APOBEC mutational signature (2014). However, we did not find *STAG2* and *APOBEC3B* mRNA levels to be significantly correlated in our cases (data not shown). Moreover, in our study all of the *STAG2* mutant cases we found in our exome sequencing study presented a high frequency of C>T alterations, and we did not identify differences in the frequency of C>T mutations according to *STAG2* mutational status. Additionally, the APOBEC signature is not associated to tumor grade or stage in UBC (Nordentoft et al. 2014), whereas the association of STAG2 with these parameters is very consistent across studies. Therefore, more work is needed to confirm the findings of Norderntoft and colleagues.

7. *STAG2* MUTATIONS AND ANEUPLOIDY: UNRELATED IN UBC

We were struck by the finding that STAG2 alterations and loss of expression were more frequent among non-aggressive UBCs, which are generally genomically stable. We then went on to confirm that STAG2 loss did not consistently lead to aneuploidy in UBC. Among 23 tumors of the non-aggressive category, there was no difference in the aneuploidy rate between STAG2 negative or positive tumors and 82% of STAG2-negative tumors lacked aneuploidy. Additionally, knocking down STAG2 in 3 different UBC cells failed to yield consistent, significant changes in chromosome number of metaphase-arrested cells. The two groups reporting an important role of STAG2 in UBC concurrently with us reached the opposite conclusion (Guo et al. 2013; Solomon et al. 2013). However, we believe that their conclusion is not supported by their data which - indeed - support our findings. Guo et al. reported that *STAG2*-mutant tumors exhibited a significantly higher genomic instability score than *STAG2* wild type tumors (2013). However, genomic instability does not equal aneuploidy (Gordon et al. 2012) and - in fact - their study does not find significant differences when considering whole chromosome changes. Solomon et al. described that 9/12 (75%) *STAG2* mutant and 10/12 (83%) *STAG2* wild-type cases exhibited chromosomal aberrations (2013). Importantly, they inappropriately defined aneuploidy as clonal chromosomal aberrations rather than whole chromosome copy number changes. Their study, as ours, found *STAG2* mutant cases lacking aneuploidy. Moreover, none of these studies performed a stratified analysis comparing tumors in the aggressive vs. non-aggressive categories. Their results possibly reflect the fact that about half of the *STAG2* mutant and wild-type cases they analyzed are aggressive, and aggressive UBCs are frequently aneuploid. Moreover, they did not find changes in chromosome counts upon introduction of wild-type *STAG2* in a non-expressing UBC cell line.

Finally, a more recent study including a cohort of 220 cases reported an inverse correlation between mutations in STAG2 and both whole chromosome copy number changes and fraction of genome altered (Taylor et al. 2014). Therefore, we believe that our conclusion that STAG2 loss does not lead on increased aneuploidy is supported by the available data. The findings in UBC are consistent with observations in myeloid malignancies where *STAG2* mutations have been reported to occur predominantly in the subgroup of diploid tumors (Welch et al. 2012; Kon et al. 2013). Additionally, a large study of Ewing's sarcoma cases confirmed frequent mutations of *STGA2* in a cohort exhibiting a very low rate of aneuploidy (Brohl et al. 2014).

We do not imply that *STAG2* alterations are universally unrelated to aneuploidy. It is possible that the functions of STAG2/cohesin, and the effects of *STAG2* inactivation, are tissue-specific.

Chromosomal instability is a frequent event in colorectal cancer (CRC) and STAG2 knockdown in CRC cells has been shown to increase chromosome numbers and to cause centromeric separation of sister chromatids (Barber et al. 2008). Similarly, Solomon and colleagues reported an increase in chromosome number upon *STAG2* silencing in a near-diploid CRC cell line, as well as a decrease in the average number of chromosomes upon STAG2 reintroduction in a mutant glioblastoma line (2011).

Given all the data available at the moment, we hypothesize that STAG2 acts through mechanisms different from its role in chromosome segregation in non-aggressive UBC development. A caveat to the functional studies reported to date is that they have been performed predominantly using established cancer cell lines. There is a need to more carefully determine the effect of *STAG2* inactivation on chromosome segregation in normal cells *in vitro* and *in vivo*.

8. STAG2 IN UBC: WHAT, IF NOT COHESION?

The findings summarized above raise the question of the mechanism through which STAG2 loss contributes to UBC. Besides its role in chromosome segregation, the cohesin complex has been shown to influence the three-dimensional structure of DNA, regulating gene transcription (reviewed in Losada 2014), and to participate in DNA repair. Genome-wide, the distribution of cohesin largely overlaps with that of CTCF in humans and mice (Parelho et al. 2008; Wendt et al. 2008), regardless of whether the complex contains STAG1 (Rubio et al. 2008) or STAG2 (Xiao et al. 2011). In MEFs, the transcriptional regulatory role of cohesin was strictly dependent on Stag1 and Stag2 could not replace it in this function (Remeseiro et al. 2012b). In agreement with this observation, depleting STAG2 or re-expressing it in wild-type or mutant human glioblastoma, CRC, and neuroglioma cell lines did not significantly alter their transcriptional profile assessed by microarrays (Solomon et al. 2011). To explore a role in gene regulation in the urothelium, we analyzed the transcriptome of control and STAG2-interfered UBC cells using RNA-sequencing which has the power of detecting low-abundance transcripts - unlike microarrays (Malone and Oliver 2011). However, we found less than 1% of genes being differentially expressed with no genes found to be commonly altered in two different cell lines. In this experiment, STAG2 downregulation was efficient and we were able to validate changes in individual gene expression by RT-qPCR, proving that the sequencing results were reliable. These exploratory results do not provide support to the notion that STAG2 is a main regulator of gene expression in UBC. However, there are several caveats to this preliminary conclusion including that the experiment

was performed using established tumor cell lines which might not recapitulate the effects of *STAG2* inactivation in normal tissue, that the effects might not be mediated through mRNA regulation, or that residual STAG2 levels might be sufficient for its function in UBC cells. Further work is required, and is ongoing in our laboratory, to conclusively establish whether STAG2 participates in urothelial homeostasis through the regulation of transcription, including an analysis of the transcriptome of normal urothelial cells upon STAG2 knockdown *in vitro* or its *in vivo* inactivation using a Stag2 conditional mouse model.

Interestingly, knocking down STAG2 on wild-type UBC cell lines led to reduced cell proliferation, unlike we would have expected for a tumor suppressor. These results were similar to our findings with ARID1A. It is conceivable that the effects of STAG2 inactivation are context-specific. Alternatively, reduced cell proliferation upon STAG2 loss might impose the selection of genetic alterations leading to an escape phenotype associated with tumor development. These possibilities are also currently being explored in our laboratory.

The strong association of STAG2 loss and low grade, low stage, papillary tumors raises the question of the cell type where these tumors are initiated. There is some debate as to urothelial $p63^-$ umbrella and $p63^+$ basal/intermediate cells represent independent lineages (Signoretti et al. 2005). An analysis of the transcriptome of microdissected basal and umbrella cells with transcriptional signatures of stage/grade-specific UBC has led to propose that NMIUBC and MIUBC originate from different progenitors (Dancik et al. 2014). While umbrella cells do not proliferate, a subgroup of basal cells are capable of dividing (Kurzrock et al. 2008). These findings lead us to speculate that alterations of *STAG2* in umbrella cells might induce proliferation, contributing to the formation of NMIUBC, whereas MIUBC could arise from *STAG2* mutations in basal cells or from the progression of *STAG2*-altered NMI tumors. To confirm this hypothesis, it will be important to delete *STAG2* in different urothelial layers using cell-specific promoters.

Cohesin has also been shown to participate in homologous recombination-mediated repair of double strand DNA breaks. Lack of chromosome arm cohesion would hamper the use of sister chromatids as repair templates, promoting the use of the homologous chromosomes instead and thus leading to loss of heterozygosity (Stark and Jasin 2003). Even the least aggressive bladder tumors exhibit frequent LOH at chromosome 9 (Spruck et al. 1994; Hartmann et al. 2002; Mhawech-Fauceglia et al. 2006). It is possible that structural characteristics of chromosome 9 make it more prone to suffering double strand breaks than the rest of chromosomes, and that inactivation of *STAG2* in the bladder promotes LOH at this particular chromosome due to faulty

repair. This is supported by the finding that, in *Saccharomyces cerevisiae*, a reduction of 70% in cohesin levels causes defects in homologous recombination repair, but reductions in cohesion levels greater than 85% are required for chromosome segregation defects (Heidinger-Pauli et al. 2010). Upon *STAG2* inactivating mutations, STAG1-cohesin can still be formed, so it is plausible that the same effect observed in yeast is taking place in the human urothelium, with the remaining STAG1-cohesin being sufficient for appropriate chromosome segregation but not for cohesin-dependent homologous recombination.

9. CLINICAL IMPLICATIONS

A recent study integrated expression information for 12 human tumor types, finding that bladder tumors primarily fall in as many as 3 out of the 11 integrated molecular subtypes identified (Hoadley et al. 2014). This highlights the heterogeneity of UBC compared to other human tumors. Interestingly, Hoadley et al. found that all UBCs with *STAG2* mutations belong to the bladder-only subtype, enriched in *FGFR3, MLL, MLL2, MLL3, EP300*, and *KDM6* mutations, among others. Additionally, UBCs with *ARID1A* mutations were also enriched in that same subtype. This highlights the importance of these genes in the urothelial context, the clinical implications of which we will explore here.

In other tumor types (ovarian clear cell carcinoma, endometrium, breast, cervix), where more studies have been carried out, ARID1A expression has not been found to be consistently associated with prognosis (Maeda et al. 2010; Katagiri et al. 2012; Lowery et al. 2012; Mamo et al. 2012; Cho et al. 2013; Fadare et al. 2013; Rahman et al. 2013; Zhang et al. 2013; Allo et al. 2014). On the other hand, multivariate analyses in clear cell renal tumors (Lichner et al. 2013) and a subset of gastric tumors (Abe et al. 2012) found significant, independent associations between loss of ARID1A and poor prognosis.

In UBC, the studies of Gui, the CGARN, and Ross did not assess the relationship between ARID1A loss or alterations and patient prognosis (2011; 2014; 2014a). Nordentoft and colleagues observed a non-significant tendency for ARID1A-negative MIUBCs to exhibit increased progression free survival in a univariate analysis (2014). Finally, a recent study focused on high-grade UBCs did not identify a significant relationship between *ARID1A* mutations and patient outcome (Kim et al. 2014). In our series, comprising a broader disease spectrum, we have shown that UBC patients exhibiting low levels of ARID1A are more likely to experience tumor progression and UBC-derived death. However, ARID1A expression was not an independent predictor of outcome and this association could merely reflect the fact that losses are more common among aggressive tumors.

To ascertain the prognostic significance of ARID1A loss of expression in MI tumors, larger studies allowing stratified analysis need to be performed.

Independently of their prognosis, ARID1A negative tumors could be considered as therapeutic targets. Drugs targeting epigenetic modifications might have therapeutic potential in ARID1A negative tumors. There are two DNA demethylators approved by the American Food and Drug Administration to treat myeloid malignancies: 5-Azacytidine and 5-Aza-2-deoxycytidine (reviewed in Yang et al. 2010). They are analogues of cytosine that incorporate into replicating DNA, causing global demethylation (reviewed in Kelly et al. 2010). However, these drugs are not stable in plasma and cause neutropenia (low neutrophil blood counts) at high doses, so their efficacy treating solid tumors has been shown to be limited (Pohlmann et al. 2002; Samlowski et al. 2005). Nevertheless, a recent preclinical trial has shown that dogs with naturally occurring MIUBC respond favorably to 5-Azacytidine treatment (Hahn et al. 2012). It would be informative to compare the effect of 5-Azacytidine in subjects with and without alterations in members of the SWI/SNF complex, including ARID1A in both clinical and pre-clinical trials.

As reviewed by Sajesh and colleagues, synthetic lethality therapies specifically targeting tumor cell features are starting to be considered to improve cancer drug selectivity (2013). UBCs with alterations in *ARID1A* might respond to drugs targeting other pathways involved in the maintenance of genomic stability. Recently, *ARID1B* has been identified as a good candidate target gene for synthetic lethality with *ARID1A*. Depleting *ARID1B* in *ARID1A* mutated cells prevents SWI/SNF complex formation, which impairs proliferation in human tumor cells and MEFs (Helming et al. 2014).

Regarding STAG2, we describe that loss of protein expression, determined using IHC, is associated with improved prognosis in UBC, although we only found it to be an independent predictor of increased survival in MI tumors. This finding is counterintuitive considering that cohesion defects and aneuploidy might be expected to associate with more aggressive tumors, but it is consistent with the strong association of *STAG2* inactivation and tumors of lower stage/grade. Solomon et al showed that loss of STAG2 - assessed using IHC - was significantly correlated with increased survival in NMIUBC. However, they found that it was associated with increased recurrence and mortality in MIUBC (Solomon et al. 2013). Guo et al. described that patients with STAG2 mutations presented significantly lower survival in both NMI and MI UBC (2013). However, these authors did not test for IHC expression and included in the analysis missense mutations whose biological significance is not known.

Several factors could explain the observed discrepancies. Our larger tumor series allowed us to perform both stratified and multivariate analysis, which strengthens the statistical power of our findings. In contrast, both Solomon and colleagues and Guo and colleagues performed only Kaplan-Meier analyses, not considering the effect of other risk factors such as tumor stage and grade (2013; 2013). The strong association between *STAG2* alterations and stage and grade makes the use of multivariable analysis a must to assess whether STAG2 expression studies might be of any clinical usefulness. Additionally, Solomon et al. did not specify how they defined loss of STAG2 expression (2013); differences in the scoring systems could account for the observed disparities. Finally, the study of Guo et al. had a very small sample size (2013). Additional, carefully stratified, analyses accounting for these potential confounders are required to draw definite conclusions about the putative clinical usefulness of the study of *STAG2* alterations in UBC.

The study of *STAG2* alterations might unveil novel therapeutic opportunities for UBC. A recent study in pancreatic cancer cell lines showed that loss of STAG2 increased sensitivity to platinum-based treatments (specifically, to carboplatin, cisplatin, and oxaliplatin) (Evers et al. 2014). Cisplatin is the standard for MIUBC treatment (Stenzl et al. 2011); therefore it will be important to determine whether *STAG2* status could predict the efficiency of chemotherapy treatment. If so, this might explain the improved outcome we observed in MIUBC patients whose tumors displayed loss of STAG2. One possible mechanism for this increase in sensitivity is that STAG2 negative tumors are especially sensitive to agents inducing DNA damage because of the role of cohesion in double-strand break repair mediated by homologous recombination.

In agreement with this potential mechanism, studies performed in yeast and later confirmed in *Caenorhabditis elegans* have suggested the existence of a synthetic lethal relationship between mutations in cohesin complex members and in *PARP1* (McLellan et al. 2012). In human CRC cells, this synthetic lethality was shown for *PARP1* and either *STAG1* or *STAG3* (O'Neil et al. 2013). A recent study showed that loss of STAG2 causes hypersensitivity to PARP inhibitors in glioblastoma cell lines (Bailey et al. 2014). Moreover, the effect of agents inhibiting PARP was maximized upon treatment with drugs causing DNA damage. However, none of the 9 different PARP inhibitors currently undergoing clinical trials focuses on UBC. Preliminary experiments in UBC cell lines did not show increased mortality upon STAG2 silencing accompanied by PARPi treatment (data not shown). However, the development of the more physiological UBC STAG2 knockout models in our laboratory will be paramount to evaluate this therapeutic strategy in bladder tumors *in vivo*.

Conclusions

In light of the results presented here, the conclusions drawn from the study are:

1. Pathways frequently altered in UBC include chromatin remodeling, DNA repair, and chromosome segregation. We have identified recurrent mutations in genes for which a role in UBC had not been previously reported: *MLL2*, *ASXL2*, *BPTF* (chromatin remodeling); *ATM*, *ERCC2*, *FANCA* (DNA repair); and *STAG1*, *STAG2*, *SMC1A*, *SMC1*β (chromosome segregation).

2. UBCs harbor an average of 169 somatic exomic mutations, with a predominance of synonymous mutations. The most common nucleotide changes are C>T/G>A transitions. We have not found significant associations between mutation number or type and age, tumor aggressiveness, or patient smoking status.

3. *ARID1A* truncating mutations frequently lead to loss of protein expression. Loss of ARID1A is commonly accompanied by low levels of FGFR3, and KRT20. ARID1A mutations and loss of expression are more common in aggressive than in non-aggressive tumors and associate with an increased risk of tumor progression.

4. *STAG2* truncating mutations frequently lead to loss of protein expression and are more common in non-aggressive than in aggressive tumors. Loss of STAG2 expression significantly associates with unicentric, small, NMI tumors of low stage and grade. STAG2 loss associates with low proliferative index, mutations in *FGFR3*, and normal p53 expression. In MIUBC, loss of STAG2 expression is an independent predictor of survival.

5. In addition to point mutations, genomic losses contribute to STAG2 inactivation.

6. Loss of STAG2 expression does not associate with increased aneuploidy in UBC tumors or cell lines, nor with increased centromeric cohesion defects *in vitro*, supporting the notion that STAG2 inactivation contributes to UBC through mechanisms different from those involved in chromosome segregation.

Conclusiones

A la luz de los resultados presentados, las conclusiones obtenidas en este estudio son:

1. Las vías alteradas frecuentemente en el CV incluyen remodelación de la cromatina, reparación del ADN y segregación cromosómica. Hemos identificado mutaciones recurrentes en genes cuya función en el CV no había sido descrita con anterioridad: *MML2*, *ASXL2* y *BPTF* (remodelación de la cromatina; *ATM*, *ERCC2* y *FANCA* (reparación del ADN); y *STAG1*, *STAG2*, *SMC1A* y *SMC1*β (segregación cromosómica).

2. Los CV presentan una media de 169 mutaciones somáticas en el exoma, con un 88% de las muestras que exhiben un ratio de mutaciones no sinónimas vs. sinónimas <1. Los cambios de nucleótidos más comunes son las transiciones C>T/G>A. No hay diferencias significativas en número o tipo de mutación dependiendo de la agresividad del tumor, la edad del paciente, ni de su exposición al tabaco.

3. Las mutaciones truncantes en *ARID1A* son más comunes en los tumores agresivos que en los no agresivos y suelen conllevar la pérdida de expresión de la proteína. La pérdida de expresión de ARID1A suele ir acompañada de niveles bajos de FGFR3 y KRT20 y, en un subgrupo de casos, también de niveles altos de KRT5/6 y KRT14. La pérdida de ARID1A se correlaciona con tasas más bajas de recurrencia tumoral y con un incremento en el riesgo de progresión tumoral.

4. Las mutaciones truncantes en STAG2 son más comunes en los tumores no agresivos que en los agresivos y suelen conllevar la pérdida de expresión de la proteína. La pérdida de expresión de STAG2 se asocia significativamente con tumores no invasivos de bajo estadio y grado que son monocéntricos y de menor tamaño. La pérdida de STAG2 se asocia con un bajo índice proliferativo, mutaciones en *FGFR3* y expresión normal de p53. En los CV invasivos, la pérdida de expresión de STAG2 es un predictor independiente de la supervivencia.

5. Además de las mutaciones puntuales, hay pérdidas genómicas que contribuyen a la inactivación de STAG2.

6. La pérdida de expresión de STAG2 no se asocia con un incremento en la aneuploidía en los tumores o líneas celulares de CV ni con mayor número de defectos en la cohesión centromérica *in vitro*, apoyando la noción de que la inactivación de STAG2 contribuye al CV a través de mecanismos distintos a los involucrados en segregación cromosómica.

References

Abe, H., D. Maeda, R. Hino, Y. Otake, M. Isogai, A. S. Ushiku, K. Matsusaka, et al. 2012. ARID1A expression loss in gastric cancer: Pathway-dependent roles with and without epstein-barr virus infection and microsatellite instability. *Virchows Archiv : An International Journal of Pathology* 461 (4) (Oct): 367-77.

Adar, R., E. Monsonego-Ornan, P. David, and A. Yayon. 2002. Differential activation of cysteine-substitution mutants of fibroblast growth factor receptor 3 is determined by cysteine localization. *Journal of Bone and Mineral Research : The Official Journal of the American Society for Bone and Mineral Research* 17 (5) (May): 860-8.

Alexandrov, L. B., S. Nik-Zainal, D. C. Wedge, S. A. Aparicio, S. Behjati, A. V. Biankin, G. R. Bignell, et al. 2013. Signatures of mutational processes in human cancer. *Nature* 500 (7463) (Aug 22): 415-21.

Allo, G., M. Q. Bernardini, R. C. Wu, IeM Shih, S. Kalloger, A. Pollett, C. B. Gilks, and B. A. Clarke. 2014. ARID1A loss correlates with mismatch repair deficiency and intact p53 expression in high-grade endometrial carcinomas. *Modern Pathology : An Official Journal of the United States and Canadian Academy of Pathology, Inc* 27 (2) (Feb): 255-61.

Allory, Y., W. Beukers, A. Sagrera, M. Flandez, M. Marques, M. Marquez, K. A. van der Keur, et al. 2014. Telomerase reverse transcriptase promoter mutations in bladder cancer: High frequency across stages, detection in urine, and lack of association with outcome. *European Urology* 65 (2) (Feb): 360-6.

Amin, M. B., J. K. McKenney, G. P. Paner, D. E. Hansel, D. J. Grignon, R. Montironi, O. Lin, et al. 2013. ICUD-EAU international consultation on bladder cancer 2012: Pathology. *European Urology* 63 (1) (Jan): 16-35.

Apodaca, G. 2004. The uroepithelium: Not just a passive barrier. *Traffic (Copenhagen, Denmark)* 5 (3) (Mar): 117-28.

Aveyard, J. S., A. Skilleter, T. Habuchi, and M. A. Knowles. 1999. Somatic mutation of PTEN in bladder carcinoma. *British Journal of Cancer* 80 (5-6) (May): 904-8.

Avritscher, E. B., C. D. Cooksley, H. B. Grossman, A. L. Sabichi, L. Hamblin, C. P. Dinney, and L. S. Elting. 2006. Clinical model of lifetime cost of treating bladder cancer and associated complications. *Urology* 68 (3) (Sep): 549-53.

Ayhan, A., T. L. Mao, T. Seckin, C. H. Wu, B. Guan, H. Ogawa, M. Futagami, et al. 2012. Loss of ARID1A expression is an early molecular event in tumor progression from ovarian endometriotic cyst to clear cell and endometrioid carcinoma. *International Journal of Gynecological Cancer : Official Journal of the International Gynecological Cancer Society* 22 (8) (Oct): 1310-5.

Babjuk, M., W. Oosterlinck, R. Sylvester, E. Kaasinen, A. Bohle, J. Palou-Redorta, M. Roupret, and European Association of Urology (EAU). 2011. EAU guidelines on non-muscle-invasive urothelial carcinoma of the bladder, the 2011 update. *European Urology* 59 (6) (Jun): 997-1008.

Bader, A. G., S. Kang, and P. K. Vogt. 2006. Cancer-specific mutations in PIK3CA are oncogenic in vivo. *Proceedings of the National Academy of Sciences of the United States of America* 103 (5) (Jan 31): 1475-9.

Bader, A. G., S. Kang, L. Zhao, and P. K. Vogt. 2005. Oncogenic PI3K deregulates transcription and translation. *Nature Reviews.Cancer* 5 (12) (Dec): 921-9.

Bailey, M. L., N. J. O'Neil, D. M. van Pel, D. A. Solomon, T. Waldman, and P. Hieter. 2014. Glioblastoma cells containing mutations in the cohesin component STAG2 are sensitive to PARP inhibition. *Molecular Cancer Therapeutics* 13 (3) (Mar): 724-32.

Banerji, S., K. Cibulskis, C. Rangel-Escareno, K. K. Brown, S. L. Carter, A. M. Frederick, M. S. Lawrence, et al. 2012. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* 486 (7403) (Jun 20): 405-9.

Barber, T. D., K. McManus, K. W. Yuen, M. Reis, G. Parmigiani, D. Shen, I. Barrett, et al. 2008. Chromatid cohesion defects may underlie chromosome instability in human colorectal

cancers. *Proceedings of the National Academy of Sciences of the United States of America* 105 (9) (Mar 4): 3443-8.

Barbieri, C. E., S. C. Baca, M. S. Lawrence, F. Demichelis, M. Blattner, J. P. Theurillat, T. A. White, et al. 2012. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nature Genetics* 44 (6) (May 20): 685-9.

Baylin, S. B., and P. A. Jones. 2011. A decade of exploring the cancer epigenome - biological and translational implications. *Nature Reviews.Cancer* 11 (10) (Sep 23): 726-34.

Bellmunt, J., B. T. Teh, G. Tortora, and J. E. Rosenberg. 2013. Molecular targets on the horizon for kidney and urothelial cancer. *Nature Reviews.Clinical Oncology* 10 (10) (Oct): 557-70.

Beothe, T., A. Nagy, L. Farkas, and G. Kovacs. 2012. P53 mutation and LOH at chromosome 9 in urothelial carcinoma. *Anticancer Research* 32 (2) (Feb): 523-7.

Biankin, A. V., N. Waddell, K. S. Kassahn, M. C. Gingras, L. B. Muthuswamy, A. L. Johns, D. K. Miller, et al. 2012. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* 491 (7424) (Nov 15): 399-405.

Birkhahn, M., A. P. Mitra, A. J. Williams, G. Lam, W. Ye, R. H. Datar, M. Balic, S. Groshen, K. E. Steven, and R. J. Cote. 2010. Predicting recurrence and progression of noninvasive papillary bladder cancer at initial presentation based on quantitative gene expression profiles. *European Urology* 57 (1) (Jan): 12-20.

Blaveri, E., J. L. Brewer, R. Roydasgupta, J. Fridlyand, S. DeVries, T. Koppie, S. Pejavar, et al. 2005. Bladder cancer stage and outcome by array-based comparative genomic hybridization. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research* 11 (19 Pt 1) (Oct 1): 7012-22.

Bosse, T., N. T. ter Haar, L. M. Seeber, P. J. v Diest, F. J. Hes, H. F. Vasen, R. A. Nout, C. L. Creutzberg, H. Morreau, and V. T. Smit. 2013. Loss of ARID1A expression and its relationship with PI3K-akt pathway alterations, TP53 and microsatellite instability in endometrial cancer. *Modern Pathology : An Official Journal of the United States and Canadian Academy of Pathology, Inc* 26 (11) (Nov): 1525-35.

Brennan, C. W., R. G. Verhaak, A. McKenna, B. Campos, H. Noushmehr, S. R. Salama, S. Zheng, et al. 2013. The somatic genomic landscape of glioblastoma. *Cell* 155 (2) (Oct 10): 462-77.

Brennan, P., O. Bogillot, S. Cordier, E. Greiser, W. Schill, P. Vineis, G. Lopez-Abente, et al. 2000. Cigarette smoking and bladder cancer in men: A pooled analysis of 11 case-control studies. *International Journal of Cancer.Journal International Du Cancer* 86 (2) (Apr 15): 289-94.

Brohl, A. S., D. A. Solomon, W. Chang, J. Wang, Y. Song, S. Sindiri, R. Patidar, et al. 2014. The genomic landscape of the ewing sarcoma family of tumors reveals recurrent STAG2 mutation. *PLoS Genetics* 10 (7) (Jul 10): e1004475.

Bulashevska, S., O. Szakacs, B. Brors, R. Eils, and G. Kovacs. 2004. Pathways of urothelial cancer progression suggested by bayesian network analysis of allelotyping data. *International Journal of Cancer.Journal International Du Cancer* 110 (6) (Jul 20): 850-6.

Burger, M., J. W. Catto, G. Dalbagni, H. B. Grossman, H. Herr, P. Karakiewicz, W. Kassouf, et al. 2013. Epidemiology and risk factors of urothelial bladder cancer. *European Urology* 63 (2) (Feb): 234-41.

Burns, M. B., N. A. Temiz, and R. S. Harris. 2013. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nature Genetics* 45 (9) (Sep): 977-83.

Cairns, P., E. Evron, K. Okami, N. Halachmi, M. Esteller, J. G. Herman, S. Bose, S. I. Wang, R. Parsons, and D. Sidransky. 1998. Point mutation and homozygous deletion of PTEN/MMAC1 in primary bladder cancers. *Oncogene* 16 (24) (Jun 18): 3215-8.

Cairns, P., A. J. Proctor, and M. A. Knowles. 1991. Loss of heterozygosity at the RB locus is frequent and correlates with muscle invasion in bladder carcinoma. *Oncogene* 6 (12) (Dec): 2305-9.

Canudas, S., and S. Smith. 2009. Differential regulation of telomere and centromere cohesion by the Scc3 homologues SA1 and SA2, respectively, in human cells. *The Journal of Cell Biology* 187 (2) (Oct 19): 165-73.

Cappellen, D., C. De Oliveira, D. Ricol, S. de Medina, J. Bourdin, X. Sastre-Garau, D. Chopin, J. P. Thiery, and F. Radvanyi. 1999. Frequent activating mutations of FGFR3 in human bladder and cervix carcinomas. *Nature Genetics* 23 (1) (Sep): 18-20.

Carro, A., D. Rico, O. M. Rueda, R. Diaz-Uriarte, and D. G. Pisano. 2010. waviCGH: A web application for the analysis and visualization of genomic copy number alterations. *Nucleic Acids Research* 38 (Web Server issue) (Jul): W182-7.

Castillo-Martin, M., J. Domingo-Domenech, O. Karni-Schmidt, T. Matos, and C. Cordon-Cardo. 2010. Molecular pathways of urothelial development and bladder tumorigenesis. *Urologic Oncology* 28 (4) (Jul-Aug): 401-8.

Cazier, J. B., S. R. Rao, C. M. McLean, A. L. Walker, B. J. Wright, E. E. Jaeger, C. Kartsonaki, et al. 2014. Whole-genome sequencing of bladder cancers reveals somatic CDKN1A mutations and clinicopathological associations with mutation burden. *Nature Communications* 5 (Apr 29): 3756.

[CGARN] Cancer Genome Atlas Network. 2012. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487 (7407) (Jul 18): 330-7.

[CGARN] Cancer Genome Atlas Research Network. 2014a. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature*.

[CGARN] Cancer Genome Atlas Research Network. 2014b. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511 (7511) (Jul 31): 543-50.

[CGARN] Cancer Genome Atlas Research Network, J. N. Weinstein, E. A. Collisson, G. B. Mills, K. R. Shaw, B. A. Ozenberger, K. Ellrott, I. Shmulevich, C. Sander, and J. M. Stuart. 2013. The cancer genome atlas pan-cancer analysis project. *Nature Genetics* 45 (10) (Oct): 1113-20.

Chan, M. W., L. W. Chan, N. L. Tang, J. H. Tong, K. W. Lo, T. L. Lee, H. Y. Cheung, et al. 2002. Hypermethylation of multiple genes in tumor tissues and voided urine in urinary bladder cancer patients. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research* 8 (2) (Feb): 464-70.

Chandler, R. L., J. Brennan, J. C. Schisler, D. Serber, C. Patterson, and T. Magnuson. 2013. ARID1a-DNA interactions are required for promoter occupancy by SWI/SNF. *Molecular and Cellular Biology* 33 (2) (Jan): 265-80.

Chan-On, W., M. L. Nairismagi, C. K. Ong, W. K. Lim, S. Dima, C. Pairojkul, K. H. Lim, et al. 2013. Exome sequencing identifies distinct mutational patterns in liver fluke-related and non-infection-related bile duct cancers. *Nature Genetics* 45 (12) (Dec): 1474-8.

Chatterjee, S. J., R. Datar, D. Youssefzadeh, B. George, P. J. Goebell, J. P. Stein, L. Young, et al. 2004a. Combined effects of p53, p21, and pRb expression in the progression of bladder transitional cell carcinoma. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* 22 (6) (Mar 15): 1007-13.

Chatterjee, S. J., B. George, P. J. Goebell, M. Alavi-Tafreshi, S. R. Shi, Y. K. Fung, P. A. Jones, C. Cordon-Cardo, R. H. Datar, and R. J. Cote. 2004b. Hyperphosphorylation of pRb: A mechanism for RB tumour suppressor pathway inactivation in bladder cancer. *The Journal of Pathology* 203 (3) (Jul): 762-70.

Cheng, L., J. C. Cheville, R. M. Neumann, B. C. Leibovich, K. S. Egan, B. E. Spotts, and D. G. Bostwick. 1999. Survival of patients with carcinoma in situ of the urinary bladder. *Cancer* 85 (11) (Jun 1): 2469-74.

Cheng, L., R. Montironi, D. D. Davidson, and A. Lopez-Beltran. 2009. Staging and reporting of urothelial carcinoma of the urinary bladder. *Modern Pathology : An Official Journal of the United States and Canadian Academy of Pathology, Inc* 22 Suppl 2 (Jun): S70-95.

Cho, H., J. S. Kim, H. Chung, C. Perry, H. Lee, and J. H. Kim. 2013. Loss of ARID1A/BAF250a expression is linked to tumor progression and adverse prognosis in cervical cancer. *Human Pathology* 44 (7) (Jul): 1365-74.

Chou, R., and T. Dana. 2010. Screening adults for bladder cancer: A review of the evidence for the U.S. preventive services task force. *Annals of Internal Medicine* 153 (7) (Oct 5): 461-8.

Cook, M. B., K. A. McGlynn, S. S. Devesa, N. D. Freedman, and W. F. Anderson. 2011. Sex disparities in cancer mortality and survival. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 20 (8) (Aug): 1629-37.

Cote, R. J., M. D. Dunn, S. J. Chatterjee, J. P. Stein, S. R. Shi, Q. C. Tran, S. X. Hu, et al. 1998. Elevated and absent pRb expression is associated with bladder cancer progression and has cooperative effects with p53. *Cancer Research* 58 (6) (Mar 15): 1090-4.

Dancik, G. M., C. R. Owens, K. A. Iczkowski, and D. Theodorescu. 2014. A cell of origin gene signature indicates human bladder cancer has distinct cellular progenitors. *Stem Cells (Dayton, Ohio)* 32 (4) (Apr): 974-82.

De Stefani, E., P. Correa, L. Fierro, E. Fontham, V. Chen, and D. Zavala. 1991. Black tobacco, mate, and bladder cancer. A case-control study from uruguay. *Cancer* 67 (2) (Jan 15): 536-40.

di Martino, E., C. G. L'Hote, W. Kennedy, D. C. Tomlinson, and M. A. Knowles. 2009. Mutant fibroblast growth factor receptor 3 induces intracellular signaling and cellular transformation in a cell type- and mutation-specific manner. *Oncogene* 28 (48) (Dec 3): 4306-16.

Duymich, C. E., J. Charlet, G. Liang, and P. A. Jones. 2013. Role of chromatin remodeler mutations in urothelial cell carcinoma of the bladder**.** Abstract. *Proceedings of the AACR Special Conference on Chromatin and Epigenetics in Cancer; Jun 19-22, 2013; Atlanta, GA. Philadelphia (PA): AACR; Cancer Res* 73(13 Suppl), no. Abstract nr A11.

Dyrskjot, L., T. Thykjaer, M. Kruhoffer, J. L. Jensen, N. Marcussen, S. Hamilton-Dutoit, H. Wolf, and T. F. Orntoft. 2003. Identifying distinct classes of bladder carcinoma using microarrays. *Nature Genetics* 33 (1) (Jan): 90-6.

Edge, S. B., and American Joint Committee on Cancer. 2010. *AJCC cancer staging manual*. New York, NY: Springer.

Engelman, J. A. 2009. Targeting PI3K signalling in cancer: Opportunities, challenges and limitations. *Nature Reviews.Cancer* 9 (8) (Aug): 550-62.

Esrig, D., D. Elmajian, S. Groshen, J. A. Freeman, J. P. Stein, S. C. Chen, P. W. Nichols, D. G. Skinner, P. A. Jones, and R. J. Cote. 1994. Accumulation of nuclear p53 and tumor progression in bladder cancer. *The New England Journal of Medicine* 331 (19) (Nov 10): 1259-64.

Euskirchen, G. M., R. K. Auerbach, E. Davidov, T. A. Gianoulis, G. Zhong, J. Rozowsky, N. Bhardwaj, M. B. Gerstein, and M. Snyder. 2011. Diverse roles and interactions of the SWI/SNF chromatin remodeling complex revealed using global approaches. *PLoS Genetics* 7 (3) (Mar): e1002008.

Evers, L., P. A. Perez-Mancera, E. Lenkiewicz, N. Tang, D. Aust, T. Knosel, P. Rummele, et al. 2014. STAG2 is a clinically relevant tumor suppressor in pancreatic ductal adenocarcinoma. *Genome Medicine* 6 (1) (Jan 31): 9.

Fadare, O., K. Gwin, M. M. Desouki, M. A. Crispens, H. W. Jones 3rd, D. Khabele, S. X. Liang, et al. 2013. The clinicopathologic significance of p53 and BAF-250a (ARID1A) expression in clear cell carcinoma of the endometrium. *Modern Pathology : An Official Journal of the United States and Canadian Academy of Pathology, Inc* 26 (8) (Aug): 1101-10.

Fajkovic, H., J. A. Halpern, E. K. Cha, A. Bahadori, T. F. Chromecki, P. I. Karakiewicz, E. Breinl, A. S. Merseburger, and S. F. Shariat. 2011. Impact of gender on bladder cancer incidence, staging, and prognosis. *World Journal of Urology* 29 (4) (Aug): 457-63.

Ferlay, J., E. Steliarova-Foucher, J. Lortet-Tieulent, S. Rosso, J. W. Coebergh, H. Comber, D. Forman, and F. Bray. 2013. Cancer incidence and mortality patterns in europe: Estimates for

40 countries in 2012. *European Journal of Cancer (Oxford, England : 1990)* 49 (6) (Apr): 1374-403.

Fernandez, M. I., J. F. Lopez, B. Vivaldi, and F. Coz. 2012. Long-term impact of arsenic in drinking water on bladder cancer health care and mortality rates 20 years after end of exposure. *The Journal of Urology* 187 (3) (Mar): 856-61.

Fernandez-Medarde, A., and E. Santos. 2011. Ras in cancer and developmental diseases. *Genes & Cancer* 2 (3) (Mar): 344-58.

Fischbach, A., A. Rogler, R. Erber, R. Stoehr, R. Poulsom, A. Heidenreich, B. S. Schneevoigt, et al. 2014. Fibroblast growth factor receptor (FGFR) amplifications are rare events in bladder cancer. *Histopathology* (Jun 4).

Frigola, J., J. Song, C. Stirzaker, R. A. Hinshelwood, M. A. Peinado, and S. J. Clark. 2006. Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. *Nature Genetics* 38 (5) (May): 540-9.

Fu, Y. P., I. Kohaar, N. Rothman, J. Earl, J. D. Figueroa, Y. Ye, N. Malats, et al. 2012. Common genetic variants in the PSCA gene influence gene expression and bladder cancer risk. *Proceedings of the National Academy of Sciences of the United States of America* 109 (13) (Mar 27): 4974-9.

Gao, X., P. Tate, P. Hu, R. Tjian, W. C. Skarnes, and Z. Wang. 2008. ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a. *Proceedings of the National Academy of Sciences of the United States of America* 105 (18) (May 6): 6656-61.

Garcia, Z., A. Kumar, M. Marques, I. Cortes, and A. C. Carrera. 2006. Phosphoinositide 3-kinase controls early and late events in mammalian cell division. *The EMBO Journal* 25 (4) (Feb 22): 655-61.

Garcia-Closas, M., N. Malats, D. Silverman, M. Dosemeci, M. Kogevinas, D. W. Hein, A. Tardon, et al. 2005. NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: Results from the spanish bladder cancer study and meta-analyses. *Lancet* 366 (9486) (Aug 20-26): 649-59.

Garcia-Closas, M., Y. Ye, N. Rothman, J. D. Figueroa, N. Malats, C. P. Dinney, N. Chatterjee, et al. 2011. A genome-wide association study of bladder cancer identifies a new susceptibility locus within SLC14A1, a urea transporter gene on chromosome 18q12.3. *Human Molecular Genetics* 20 (21) (Nov 1): 4282-9.

Gerlinger, M., A. J. Rowan, S. Horswell, J. Larkin, D. Endesfelder, E. Gronroos, P. Martinez, et al. 2012. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *The New England Journal of Medicine* 366 (10) (Mar 8): 883-92.

Gildea, J. J., M. Herlevsen, M. A. Harding, K. M. Gulding, C. A. Moskaluk, H. F. Frierson, and D. Theodorescu. 2004. PTEN can inhibit in vitro organotypic and in vivo orthotopic invasion of human bladder cancer cells even in the absence of its lipid phosphatase activity. *Oncogene* 23 (40) (Sep 2): 6788-97.

Gonzalez-Perez, A., and N. Lopez-Bigas. 2012. Functional impact bias reveals cancer drivers. *Nucleic Acids Research* 40 (21) (Nov): e169.

Gordon, D. J., B. Resio, and D. Pellman. 2012. Causes and consequences of aneuploidy in cancer. *Nature Reviews.Genetics* 13 (3) (Jan 24): 189-203.

Grossman, H. B., M. Liebert, M. Antelo, C. P. Dinney, S. X. Hu, J. L. Palmer, and W. F. Benedict. 1998. p53 and RB expression predict progression in T1 bladder cancer. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research* 4 (4) (Apr): 829-34.

Guan, B., T. L. Mao, P. K. Panuganti, E. Kuhn, R. J. Kurman, D. Maeda, E. Chen, Y. M. Jeng, T. L. Wang, and IeM Shih. 2011a. Mutation and loss of expression of ARID1A in uterine low-grade endometrioid carcinoma. *The American Journal of Surgical Pathology* 35 (5) (May): 625-32.

Guan, B., T. L. Wang, and IeM Shih. 2011b. ARID1A, a factor that promotes formation of SWI/SNF-mediated chromatin remodeling, is a tumor suppressor in gynecologic cancers. *Cancer Research* 71 (21) (Nov 1): 6718-27.

Gui, Y., G. Guo, Y. Huang, X. Hu, A. Tang, S. Gao, R. Wu, et al. 2011. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nature Genetics* 43 (9) (Aug 7): 875-8.

Guichard, C., G. Amaddeo, S. Imbeaud, Y. Ladeiro, L. Pelletier, I. B. Maad, J. Calderaro, et al. 2012. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nature Genetics* 44 (6) (May 6): 694-8.

Guo, G., X. Sun, C. Chen, S. Wu, P. Huang, Z. Li, M. Dean, et al. 2013. Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation. *Nature Genetics* 45 (12) (Dec): 1459-63.

Gust, K. M., D. J. McConkey, S. Awrey, P. K. Hegarty, J. Qing, J. Bondaruk, A. Ashkenazi, B. Czerniak, C. P. Dinney, and P. C. Black. 2013. Fibroblast growth factor receptor 3 is a rational therapeutic target in bladder cancer. *Molecular Cancer Therapeutics* 12 (7) (Jul): 1245-54.

Hafner, C., J. M. van Oers, T. Vogt, M. Landthaler, R. Stoehr, H. Blaszyk, F. Hofstaedter, E. C. Zwarthoff, and A. Hartmann. 2006. Mosaicism of activating FGFR3 mutations in human skin causes epidermal nevi. *The Journal of Clinical Investigation* 116 (8) (Aug): 2201-7.

Hahn, N. M., P. L. Bonney, D. Dhawan, D. R. Jones, C. Balch, Z. Guo, C. Hartman-Frey, et al. 2012. Subcutaneous 5-azacitidine treatment of naturally occurring canine urothelial carcinoma: A novel epigenetic approach to human urothelial carcinoma drug development. *The Journal of Urology* 187 (1) (Jan): 302-9.

Harris, L. D., J. De La Cerda, T. Tuziak, D. Rosen, L. Xiao, Y. Shen, A. L. Sabichi, B. Czerniak, and H. B. Grossman. 2008. Analysis of the expression of biomarkers in urinary bladder cancer using a tissue microarray. *Molecular Carcinogenesis* 47 (9) (Sep): 678-85.

Hartmann, A., G. Schlake, D. Zaak, E. Hungerhuber, A. Hofstetter, F. Hofstaedter, and R. Knuechel. 2002. Occurrence of chromosome 9 and p53 alterations in multifocal dysplasia and carcinoma in situ of human urinary bladder. *Cancer Research* 62 (3) (Feb 1): 809-18.

Haugsten, E. M., A. Wiedlocha, S. Olsnes, and J. Wesche. 2010. Roles of fibroblast growth factor receptors in carcinogenesis. *Molecular Cancer Research : MCR* 8 (11) (Nov): 1439-52.

Heidinger-Pauli, J. M., O. Mert, C. Davenport, V. Guacci, and D. Koshland. 2010. Systematic reduction of cohesin differentially affects chromosome segregation, condensation, and DNA repair. *Current Biology : CB* 20 (10) (May 25): 957-63.

Helming, K. C., X. Wang, B. G. Wilson, F. Vazquez, J. R. Haswell, H. E. Manchester, Y. Kim, et al. 2014. ARID1B is a specific vulnerability in ARID1A-mutant cancers. *Nature Medicine* 20 (3) (Mar): 251-4.

Hernandez, S., E. Lopez-Knowles, J. Lloreta, M. Kogevinas, A. Amoros, A. Tardon, A. Carrato, C. Serra, N. Malats, and F. X. Real. 2006. Prospective study of FGFR3 mutations as a prognostic factor in nonmuscle invasive urothelial bladder carcinomas. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* 24 (22) (Aug 1): 3664-71.

Hernandez, S., E. Lopez-Knowles, J. Lloreta, M. Kogevinas, R. Jaramillo, A. Amoros, A. Tardon, et al. 2005. FGFR3 and Tp53 mutations in T1G3 transitional bladder carcinomas: Independent distribution and lack of association with prognosis. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research* 11 (15) (Aug 1): 5444-50.

Hitchings, A. W., M. Kumar, S. Jordan, V. Nargund, J. Martin, and D. M. Berney. 2004. Prediction of progression in pTa and pT1 bladder carcinomas with p53, p16 and pRb. *British Journal of Cancer* 91 (3) (Aug 2): 552-7.

Ho, J. S., W. Ma, D. Y. Mao, and S. Benchimol. 2005. p53-dependent transcriptional repression of c-myc is required for G1 cell cycle arrest. *Molecular and Cellular Biology* 25 (17) (Sep): 7423-31.

Hoadley, K. A., C. Yau, D. M. Wolf, A. D. Cherniack, D. Tamborero, S. Ng, M. D. Leiserson, et al. 2014. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* 158 (4) (Aug 14): 929-44.

Horn, S., A. Figl, P. S. Rachakonda, C. Fischer, A. Sucker, A. Gast, S. Kadel, et al. 2013. TERT promoter mutations in familial and sporadic melanoma. *Science (New York, N.Y.)* 339 (6122) (Feb 22): 959-61.

Hovestadt, V., D. T. Jones, S. Picelli, W. Wang, M. Kool, P. A. Northcott, M. Sultan, et al. 2014. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* 510 (7506) (Jun 26): 537-41.

Huang, F. W., E. Hodis, M. J. Xu, G. V. Kryukov, L. Chin, and L. A. Garraway. 2013. Highly recurrent TERT promoter mutations in human melanoma. *Science (New York, N.Y.)* 339 (6122) (Feb 22): 957-9.

Huang, J., Q. Deng, Q. Wang, K. Y. Li, J. H. Dai, N. Li, Z. D. Zhu, et al. 2012. Exome sequencing of hepatitis B virus-associated hepatocellular carcinoma. *Nature Genetics* 44 (10) (Oct): 1117-21.

Hurst, C. D., F. M. Platt, and M. A. Knowles. 2014. Comprehensive mutation analysis of the TERT promoter in bladder cancer and detection of mutations in voided urine. *European Urology* 65 (2) (Feb): 367-9.

Imielinski, M., A. H. Berger, P. S. Hammerman, B. Hernandez, T. J. Pugh, E. Hodis, J. Cho, et al. 2012. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 150 (6) (Sep 14): 1107-20.

Iyer, G., H. Al-Ahmadie, N. Schultz, A. J. Hanrahan, I. Ostrovnaya, A. V. Balar, P. H. Kim, et al. 2013. Prevalence and co-occurrence of actionable genomic alterations in high-grade bladder cancer. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* 31 (25) (Sep 1): 3133-40.

Jacobs, K. B., M. Yeager, W. Zhou, S. Wacholder, Z. Wang, B. Rodriguez-Santiago, A. Hutchinson, et al. 2012. Detectable clonal mosaicism and its relationship to aging and cancer. *Nature Genetics* 44 (6) (May 6): 651-8.

Jain, A. N., T. A. Tokuyasu, A. M. Snijders, R. Segraves, D. G. Albertson, and D. Pinkel. 2002. Fully automatic quantification of microarray image data. *Genome Research* 12 (2) (Feb): 325-32.

James, N. D., S. A. Hussain, E. Hall, P. Jenkins, J. Tremlett, C. Rawlings, M. Crundwell, et al. 2012. Radiotherapy with or without chemotherapy in muscle-invasive bladder cancer. *The New England Journal of Medicine* 366 (16) (Apr 19): 1477-88.

Jebar, A. H., C. D. Hurst, D. C. Tomlinson, C. Johnston, C. F. Taylor, and M. A. Knowles. 2005. FGFR3 and ras gene mutations are mutually exclusive genetic events in urothelial cell carcinoma. *Oncogene* 24 (33) (Aug 4): 5218-25.

Jemal, A., F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman. 2011. Global cancer statistics. *CA: A Cancer Journal for Clinicians* 61 (2) (Mar-Apr): 69-90.

Jones, S., T. L. Wang, IeM Shih, T. L. Mao, K. Nakayama, R. Roden, R. Glas, et al. 2010. Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science (New York, N.Y.)* 330 (6001) (Oct 8): 228-31.

Jost, S. P., J. A. Gosling, and J. S. Dixon. 1989. The morphology of normal human bladder urothelium. *Journal of Anatomy* 167 (Dec): 103-15.

Jung, I., and E. Messing. 2000. Molecular mechanisms and pathways in bladder cancer development and progression. *Cancer Control : Journal of the Moffitt Cancer Center* 7 (4) (Jul-Aug): 325-34.

Kadoch, C., D. C. Hargreaves, C. Hodges, L. Elias, L. Ho, J. Ranish, and G. R. Crabtree. 2013. Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nature Genetics* 45 (6) (Jun): 592-601.

Kamat, A. M., P. K. Hegarty, J. R. Gee, P. E. Clark, R. S. Svatek, N. Hegarty, S. F. Shariat, et al. 2013. ICUD-EAU international consultation on bladder cancer 2012: Screening, diagnosis, and molecular markers. *European Urology* 63 (1) (Jan): 4-15.

Katagiri, A., K. Nakayama, M. T. Rahman, M. Rahman, H. Katagiri, M. Ishikawa, T. Ishibashi, et al. 2012. Frequent loss of tumor suppressor ARID1A protein expression in adenocarcinomas/adenosquamous carcinomas of the uterine cervix. *International Journal of Gynecological Cancer : Official Journal of the International Gynecological Cancer Society* 22 (2) (Feb): 208-12.

Kausch, I., M. Sommerauer, F. Montorsi, A. Stenzl, D. Jacqmin, P. Jichlinski, D. Jocham, A. Ziegler, and R. Vonthein. 2010. Photodynamic diagnosis in non-muscle-invasive bladder cancer: A systematic review and cumulative analysis of prospective studies. *European Urology* 57 (4) (Apr): 595-606.

Kelly, T. K., D. D. De Carvalho, and P. A. Jones. 2010. Epigenetic modifications as therapeutic targets. *Nature Biotechnology* 28 (10) (Oct 13): 1069-78.

Kiemeney, L. A., P. Sulem, S. Besenbacher, S. H. Vermeulen, A. Sigurdsson, G. Thorleifsson, D. F. Gudbjartsson, et al. 2010. A sequence variant at 4p16.3 confers susceptibility to urinary bladder cancer. *Nature Genetics* 42 (5) (May): 415-9.

Kiemeney, L. A., S. Thorlacius, P. Sulem, F. Geller, K. K. Aben, S. N. Stacey, J. Gudmundsson, et al. 2008. Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nature Genetics* 40 (11) (Nov): 1307-12.

Kim, P. H., E. K. Cha, J. P. Sfakianos, G. Iyer, E. C. Zabor, S. N. Scott, I. Ostrovnaya, et al. 2014. Genomic predictors of survival in patients with high-grade urothelial carcinoma of the bladder. *European Urology* (Aug 1).

Kinde, I., E. Munari, S. F. Faraj, R. H. Hruban, M. Schoenberg, T. Bivalacqua, M. Allaf, et al. 2013. TERT promoter mutations occur early in urothelial neoplasia and are biomarkers of early disease and disease recurrence in urine. *Cancer Research* 73 (24) (Dec 15): 7162-7.

Kingston, R. E., C. A. Chen, and J. K. Rose. 2003. Calcium phosphate transfection. *Current Protocols in Molecular Biology / Edited by Frederick M.Ausubel ...[Et Al.]* Chapter 9 (Aug): Unit 9.1.

Kitagawa, R. 2009. Key players in chromosome segregation in caenorhabditis elegans. *Frontiers in Bioscience (Landmark Edition)* 14 (Jan 1): 1529-57.

Knowles, M. A. 2008. Molecular pathogenesis of bladder cancer. *International Journal of Clinical Oncology* 13 (4) (Aug): 287-97.

———. 2007. Role of FGFR3 in urothelial cell carcinoma: Biomarker and potential therapeutic target. *World Journal of Urology* 25 (6) (Dec): 581-93.

———. 2006. Molecular subtypes of bladder cancer: Jekyll and hyde or chalk and cheese? *Carcinogenesis* 27 (3) (Mar): 361-73.

Kogevinas, M., A. 't Mannetje, S. Cordier, U. Ranft, C. A. Gonzalez, P. Vineis, J. Chang-Claude, et al. 2003. Occupation and bladder cancer among men in western europe. *Cancer Causes & Control : CCC* 14 (10) (Dec): 907-14.

Kompier, L. C., I. Lurkin, M. N. van der Aa, B. W. van Rhijn, T. H. van der Kwast, and E. C. Zwarthoff. 2010. FGFR3, HRAS, KRAS, NRAS and PIK3CA mutations in bladder cancer and their potential as biomarkers for surveillance and therapy. *PloS One* 5 (11) (Nov 3): e13821.

Kon, A., L. Y. Shih, M. Minamino, M. Sanada, Y. Shiraishi, Y. Nagata, K. Yoshida, et al. 2013. Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. *Nature Genetics* 45 (10) (Oct): 1232-7.

Koo, S. H., K. C. Kwon, C. H. Ihm, Y. M. Jeon, J. W. Park, and C. K. Sul. 1999. Detection of genetic alterations in bladder tumors by comparative genomic hybridization and cytogenetic analysis. *Cancer Genetics and Cytogenetics* 110 (2) (Apr 15): 87-93.

Kruger, S., A. Mahnken, I. Kausch, and A. C. Feller. 2005. P16 immunoreactivity is an independent predictor of tumor progression in minimally invasive urothelial bladder carcinoma. *European Urology* 47 (4) (Apr): 463-7.

Kumar, A., T. A. White, A. P. MacKenzie, N. Clegg, C. Lee, R. F. Dumpit, I. Coleman, et al. 2011. Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. *Proceedings of the National Academy of Sciences of the United States of America* 108 (41) (Oct 11): 17087-92.

Kumar, P., S. Henikoff, and P. C. Ng. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols* 4 (7): 1073-81.

Kurzrock, E. A., D. K. Lieu, L. A. Degraffenried, C. W. Chan, and R. R. Isseroff. 2008. Label-retaining cells of the bladder: Candidate urothelial stem cells. *American Journal of Physiology.Renal Physiology* 294 (6) (Jun): F1415-21.

Laguna, P., F. Smedts, J. Nordling, T. Horn, K. Bouchelouche, A. Hopman, and J. de la Rosette. 2006. Keratin expression profiling of transitional epithelium in the painful bladder syndrome/interstitial cystitis. *American Journal of Clinical Pathology* 125 (1) (Jan): 105-10.

Lane, D. 2004. Curing cancer with p53. *The New England Journal of Medicine* 350 (26) (Jun 24): 2711-2.

Lawrence, M. S., P. Stojanov, P. Polak, G. V. Kryukov, K. Cibulskis, A. Sivachenko, S. L. Carter, et al. 2013. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499 (7457) (Jul 11): 214-8.

Lee, R. S., C. Stewart, S. L. Carter, L. Ambrogio, K. Cibulskis, C. Sougnez, M. S. Lawrence, et al. 2012. A remarkably simple genome underlies highly malignant pediatric rhabdoid cancers. *The Journal of Clinical Investigation* 122 (8) (Aug 1): 2983-8.

Le Frere-Belda, M. A., S. Gil Diez de Medina, A. Daher, N. Martin, B. Albaud, D. Heudes, C. C. Abbou, et al. 2004. Profiles of the 2 INK4a gene products, p16 and p14ARF, in human reference urothelium and bladder carcinomas, according to pRb and p53 protein status. *Human Pathology* 35 (7) (Jul): 817-24.

Le Gallo, M., A. J. O'Hara, M. L. Rudd, M. E. Urick, N. F. Hansen, N. J. O'Neil, J. C. Price, et al. 2012. Exome sequencing of serous endometrial tumors identifies recurrent somatic mutations in chromatin-remodeling and ubiquitin ligase complex genes. *Nature Genetics* 44 (12) (Dec): 1310-5.

Letasiova, S., A. Medve'ova, A. Sovcikova, M. Dusinska, K. Volkovova, C. Mosoiu, and A. Bartonova. 2012. Bladder cancer, a review of the environmental risk factors. *Environmental Health : A Global Access Science Source* 11 Suppl 1 (Jun 28): S11,069X-11-S1-S11.

Lewis, S. A. 2000. Everything you wanted to know about the bladder epithelium but were afraid to ask. *American Journal of Physiology.Renal Physiology* 278 (6) (Jun): F867-74.

Li, M., X. Fang, D. J. Baker, L. Guo, X. Gao, Z. Wei, S. Han, J. M. van Deursen, and P. Zhang. 2010. The ATM-p53 pathway suppresses aneuploidy-induced tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America* 107 (32) (Aug 10): 14188-93.

Li, W., T. Zhu, and K. L. Guan. 2004. Transformation potential of ras isoforms correlates with activation of phosphatidylinositol 3-kinase but not ERK. *The Journal of Biological Chemistry* 279 (36) (Sep 3): 37398-406.

Lichner, Z., A. Scorilas, N. M. White, A. H. Girgis, L. Rotstein, K. C. Wiegand, A. Latif, C. Chow, D. Huntsman, and G. M. Yousef. 2013. The chromatin remodeling gene ARID1A is a new prognostic marker in clear cell renal cell carcinoma. *The American Journal of Pathology* 182 (4) (Apr): 1163-70.

Lindgren, D., A. Frigyesi, S. Gudjonsson, G. Sjodahl, C. Hallden, G. Chebil, S. Veerla, et al. 2010. Combined gene expression and genomic profiling define two intrinsic molecular subtypes of urothelial carcinoma and gene signatures for molecular grading and outcome. *Cancer Research* 70 (9) (May 1): 3463-72.

Liu, X., G. Wu, Y. Shan, C. Hartmann, A. von Deimling, and M. Xing. 2013. Highly prevalent TERT promoter mutations in bladder cancer and glioblastoma. *Cell Cycle (Georgetown, Tex.)* 12 (10) (May 15): 1637-8.

Loeb, L. A., and B. D. Preston. 1986. Mutagenesis by apurinic/apyrimidinic sites. *Annual Review of Genetics* 20 : 201-30.

Lopez-Knowles, E., S. Hernandez, N. Malats, M. Kogevinas, J. Lloreta, A. Carrato, A. Tardon, C. Serra, and F. X. Real. 2006. PIK3CA mutations are an early genetic alteration associated with FGFR3 mutations in superficial papillary bladder tumors. *Cancer Research* 66 (15) (Aug 1): 7401-4.

Losada, A. 2014. Cohesin in cancer: Chromosome segregation and beyond. *Nature Reviews.Cancer* 14 (6) (Jun): 389-93.

Lowery, W. J., J. M. Schildkraut, L. Akushevich, R. Bentley, J. R. Marks, D. Huntsman, and A. Berchuck. 2012. Loss of ARID1A-associated protein expression is a frequent event in clear cell and endometrioid ovarian cancers. *International Journal of Gynecological Cancer : Official Journal of the International Gynecological Cancer Society* 22 (1) (Jan): 9-14.

Maeda, D., T. L. Mao, M. Fukayama, S. Nakagawa, T. Yano, Y. Taketani, and IeM Shih. 2010. Clinicopathological significance of loss of ARID1A immunoreactivity in ovarian clear cell carcinoma. *International Journal of Molecular Sciences* 11 (12): 5120-8.

Malats, N., A. Bustos, C. M. Nascimento, F. Fernandez, M. Rivas, D. Puente, M. Kogevinas, and F. X. Real. 2005. P53 as a prognostic marker for bladder cancer: A meta-analysis and review. *The Lancet.Oncology* 6 (9) (Sep): 678-86.

Malone, J. H., and B. Oliver. 2011. Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biology* 9 (May 31): 34,7007-9-34.

Malumbres, M., and M. Barbacid. 2009. Cell cycle, CDKs and cancer: A changing paradigm. *Nature Reviews.Cancer* 9 (3) (Mar): 153-66.

———. 2001. To cycle or not to cycle: A critical decision in cancer. *Nature Reviews.Cancer* 1 (3) (Dec): 222-31.

Mamo, A., L. Cavallone, S. Tuzmen, C. Chabot, C. Ferrario, S. Hassan, H. Edgren, et al. 2012. An integrated genomic approach identifies ARID1A as a candidate tumor-suppressor gene in breast cancer. *Oncogene* 31 (16) (Apr 19): 2090-100.

Martin, B. F. 1972. Cell replacement and differentiation in transitional epithelium: A histological and autoradiographic study of the guinea-pig bladder and ureter. *Journal of Anatomy* 112 (Pt 3) (Sep): 433-55.

Maruyama, R., S. Toyooka, K. O. Toyooka, K. Harada, A. K. Virmani, S. Zochbauer-Muller, A. J. Farinas, et al. 2001. Aberrant promoter methylation profile of bladder cancer and its relationship to clinicopathological features. *Cancer Research* 61 (24) (Dec 15): 8659-63.

McCall, M. N., B. M. Bolstad, and R. A. Irizarry. 2010. Frozen robust multiarray analysis (fRMA). *Biostatistics (Oxford, England)* 11 (2) (Apr): 242-53.

McConkey, D. J., S. Lee, W. Choi, M. Tran, T. Majewski, S. Lee, A. Siefker-Radtke, C. Dinney, and B. Czerniak. 2010. Molecular genetics of bladder cancer: Emerging mechanisms of tumor initiation and progression. *Urologic Oncology* 28 (4) (Jul-Aug): 429-40.

McLellan, J. L., N. J. O'Neil, I. Barrett, E. Ferree, D. M. van Pel, K. Ushey, P. Sipahimalani, J. Bryan, A. M. Rose, and P. Hieter. 2012. Synthetic lethality of cohesins with PARPs and replication fork mediators. *PLoS Genetics* 8 (3): e1002574.

Medina, P. P., O. A. Romero, T. Kohno, L. M. Montuenga, R. Pio, J. Yokota, and M. Sanchez-Cespedes. 2008. Frequent BRG1/SMARCA4-inactivating mutations in human lung cancer cell lines. *Human Mutation* 29 (5) (May): 617-22.

Mhawech-Fauceglia, P., R. T. Cheney, and J. Schwaller. 2006. Genetic alterations in urothelial bladder carcinoma: An updated review. *Cancer* 106 (6) (Mar 15): 1205-16.

Miñana, B., J. M. Cozar, J. Palou, M. Unda Urzaiz, R. A. Medina-Lopez, J. Subira Rios, F. de la Rosa-Kehrmann, et al. 2014. Bladder cancer in spain 2011: Population based study. *The Journal of Urology* 191 (2) (Feb): 323-8.

Miyamoto, H., T. Shuin, S. Torigoe, Y. Iwasaki, and Y. Kubota. 1995. Retinoblastoma gene mutations in primary human bladder cancer. *British Journal of Cancer* 71 (4) (Apr): 831-5.

Moll, R., D. L. Schiller, and W. W. Franke. 1990. Identification of protein IT of the intestinal cytoskeleton as a novel type I cytokeratin with unusual properties and expression patterns. *The Journal of Cell Biology* 111 (2) (Aug): 567-80.

Montironi, R., and A. Lopez-Beltran. 2005. The 2004 WHO classification of bladder tumors: A summary and commentary. *International Journal of Surgical Pathology* 13 (2) (Apr): 143-53.

Moore, L. E., R. M. Pfeiffer, C. Poscablo, F. X. Real, M. Kogevinas, D. Silverman, R. Garcia-Closas, et al. 2008. Genomic DNA hypomethylation as a biomarker for bladder cancer susceptibility in the spanish bladder cancer study: A case-control study. *The Lancet.Oncology* 9 (4) (Apr): 359-66.

Mostofi, F. K., L. H. Sobin, and H. Torloni. 1973. Histological typing of urinary tumors. In *International histological classification of tumors.*, ed. Vol. 10. Geneva: World Health Organization.

Nagl, N. G.,Jr, D. R. Zweitzig, B. Thimmapaya, G. R. Beck Jr, and E. Moran. 2006. The c-myc gene is a direct target of mammalian SWI/SNF-related complexes during differentiation-associated cell cycle arrest. *Cancer Research* 66 (3) (Feb 1): 1289-93.

Nordentoft, I., P. Lamy, K. Birkenkamp-Demtroder, K. Shumansky, S. Vang, H. Hornshoj, M. Juul, et al. 2014. Mutational context and diverse clonal development in early and late bladder cancer. *Cell Reports* 7 (5) (Jun 12): 1649-63.

Oliveros, J. C. *VENNY. An interactive tool for comparing lists with Venn diagrams*. 2007 [2014]. Available from http://bioinfogp.cnb.csic.es/tools/venny/index.html.

Olivier, M., M. Hollstein, and P. Hainaut. 2010. TP53 mutations in human cancers: Origins, consequences, and clinical use. *Cold Spring Harbor Perspectives in Biology* 2 (1) (Jan): a001008.

O'Neil, N. J., D. M. van Pel, and P. Hieter. 2013. Synthetic lethality and cancer: Cohesin and PARP at the replication fork. *Trends in Genetics : TIG* 29 (5) (May): 290-7.

Otto, W., S. Denzinger, S. Bertz, A. Gaumann, P. J. Wild, A. Hartmann, and R. Stoehr. 2009. No mutations of FGFR3 in normal urothelium in the vicinity of urothelial carcinoma of the bladder harbouring activating FGFR3 mutations in patients with bladder cancer. *International Journal of Cancer.Journal International Du Cancer* 125 (9) (Nov 1): 2205-8.

Parelho, V., S. Hadjur, M. Spivakov, M. Leleu, S. Sauer, H. C. Gregson, A. Jarmuz, et al. 2008. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132 (3) (Feb 8): 422-33.

Parkin, D. M. 2006. The global health burden of infection-associated cancers in the year 2002. *International Journal of Cancer.Journal International Du Cancer* 118 (12) (Jun 15): 3030-44.

Perini, G., D. Diolaiti, A. Porro, and G. Della Valle. 2005. In vivo transcriptional regulation of N-myc target genes is controlled by E-box methylation. *Proceedings of the National Academy of Sciences of the United States of America* 102 (34) (Aug 23): 12117-22.

Platt, F. M., C. D. Hurst, C. F. Taylor, W. M. Gregory, P. Harnden, and M. A. Knowles. 2009. Spectrum of phosphatidylinositol 3-kinase pathway gene alterations in bladder cancer. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research* 15 (19) (Oct 1): 6008-17.

Pohlmann, P., L. P. DiLeone, A. I. Cancella, A. P. Caldas, L. Dal Lago, O. Campos Jr, E. Monego, W. Rivoire, and G. Schwartsmann. 2002. Phase II trial of cisplatin plus decitabine, a new DNA hypomethylating agent, in patients with advanced squamous cell carcinoma of the cervix. *American Journal of Clinical Oncology* 25 (5) (Oct): 496-501.

Prendergast, G. C., D. Lawe, and E. B. Ziff. 1991. Association of myn, the murine homolog of max, with c-myc stimulates methylation-sensitive DNA binding and ras cotransformation. *Cell* 65 (3) (May 3): 395-407.

Prior, I. A., P. D. Lewis, and C. Mattos. 2012. A comprehensive survey of ras mutations in cancer. *Cancer Research* 72 (10) (May 15): 2457-67.

Puente, X. S., and C. Lopez-Otin. 2013. The evolutionary biography of chronic lymphocytic leukemia. *Nature Genetics* 45 (3) (Mar): 229-31.

Pugh, T. J., O. Morozova, E. F. Attiyeh, S. Asgharzadeh, J. S. Wei, D. Auclair, S. L. Carter, et al. 2013. The genetic landscape of high-risk neuroblastoma. *Nature Genetics* 45 (3) (Mar): 279-84.

Puzio-Kuter, A. M., M. Castillo-Martin, C. W. Kinkade, X. Wang, T. H. Shen, T. Matos, M. M. Shen, C. Cordon-Cardo, and C. Abate-Shen. 2009. Inactivation of p53 and pten promotes invasive bladder cancer. *Genes & Development* 23 (6) (Mar 15): 675-80.

Quesada, V., L. Conde, N. Villamor, G. R. Ordonez, P. Jares, L. Bassaganyas, A. J. Ramsay, et al. 2011. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nature Genetics* 44 (1) (Dec 11): 47-52.

Rachakonda, P. S., I. Hosen, P. J. de Verdier, M. Fallah, B. Heidenreich, C. Ryk, N. P. Wiklund, et al. 2013. TERT promoter mutations in bladder cancer affect patient survival and disease recurrence through modification by a common polymorphism. *Proceedings of the National Academy of Sciences of the United States of America* 110 (43) (Oct 22): 17426-31.

Rafnar, T., P. Sulem, S. N. Stacey, F. Geller, J. Gudmundsson, A. Sigurdsson, M. Jakobsdottir, et al. 2009. Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nature Genetics* 41 (2) (Feb): 221-7.

Rafnar, T., S. H. Vermeulen, P. Sulem, G. Thorleifsson, K. K. Aben, J. A. Witjes, A. J. Grotenhuis, et al. 2011. European genome-wide association study identifies SLC14A1 as a new urinary bladder cancer susceptibility gene. *Human Molecular Genetics* 20 (21) (Nov 1): 4268-81.

Rahman, M., K. Nakayama, M. T. Rahman, H. Katagiri, A. Katagiri, T. Ishibashi, M. Ishikawa, K. Iida, and K. Miyazaki. 2013. Clinicopathologic analysis of loss of AT-rich interactive domain 1A expression in endometrial cancer. *Human Pathology* 44 (1) (Jan): 103-9.

Real, F. X., P. C. Boutros, and N. Malats. 2014. Next-generation sequencing of urologic cancers: Next is now. *European Urology* 66 (1) (Jul): 4-7.

Rebouissou, S., A. Herault, E. Letouze, Y. Neuzillet, A. Laplanche, K. Ofualuka, P. Maille, et al. 2012. CDKN2A homozygous deletion is associated with muscle invasion in FGFR3-mutated urothelial bladder carcinoma. *The Journal of Pathology* 227 (3) (Jul): 315-24.

Remeseiro, S., A. Cuadrado, M. Carretero, P. Martinez, W. C. Drosopoulos, M. Canamero, C. L. Schildkraut, M. A. Blasco, and A. Losada. 2012a. Cohesin-SA1 deficiency drives aneuploidy and tumourigenesis in mice due to impaired replication of telomeres. *The EMBO Journal* 31 (9) (May 2): 2076-89.

Remeseiro, S., A. Cuadrado, G. Gomez-Lopez, D. G. Pisano, and A. Losada. 2012b. A unique role of cohesin-SA1 in gene regulation and development. *The EMBO Journal* 31 (9) (May 2): 2090-102.

Remeseiro, S., and A. Losada. 2013. Cohesin, a chromatin engagement ring. *Current Opinion in Cell Biology* 25 (1) (Feb): 63-71.

Reva, B., Y. Antipin, and C. Sander. 2011. Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Research* 39 (17) (Sep 1): e118.

Rhodes, J. M., M. McEwan, and J. A. Horsfield. 2011. Gene regulation by cohesin in cancer: Is the ring an unexpected party to proliferation? *Molecular Cancer Research : MCR* 9 (12) (Dec): 1587-607.

Rieger, K. M., A. F. Little, J. M. Swart, W. V. Kastrinakis, J. M. Fitzgerald, D. T. Hess, J. A. Libertino, and I. C. Summerhayes. 1995. Human bladder carcinoma cell lines as indicators of oncogenic change relevant to urothelial neoplastic progression. *British Journal of Cancer* 72 (3) (Sep): 683-90.

Rink, M., M. Babjuk, J. W. Catto, P. Jichlinski, S. F. Shariat, A. Stenzl, H. Stepp, D. Zaak, and J. A. Witjes. 2013. Hexyl aminolevulinate-guided fluorescence cystoscopy in the diagnosis and follow-up of patients with non-muscle-invasive bladder cancer: A critical review of the current literature. *European Urology* 64 (4) (Oct): 624-38.

Roberts, S. A., M. S. Lawrence, L. J. Klimczak, S. A. Grimm, D. Fargo, P. Stojanov, A. Kiezun, et al. 2013. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nature Genetics* 45 (9) (Sep): 970-6.

Rodriguez-Nieto, S., A. Canada, E. Pros, A. I. Pinto, J. Torres-Lanzas, F. Lopez-Rios, L. Sanchez-Verde, D. G. Pisano, and M. Sanchez-Cespedes. 2011. Massive parallel DNA pyrosequencing analysis of the tumor suppressor BRG1/SMARCA4 in lung primary tumors. *Human Mutation* 32 (2) (Feb): E1999-2017.

Rodriguez-Santiago, B., N. Malats, N. Rothman, L. Armengol, M. Garcia-Closas, M. Kogevinas, O. Villa, et al. 2010. Mosaic uniparental disomies and aneuploidies as large structural variants of the human genome. *American Journal of Human Genetics* 87 (1) (Jul 9): 129-38.

Romero, O. A., F. Setien, S. John, P. Gimenez-Xavier, G. Gomez-Lopez, D. Pisano, E. Condom, A. Villanueva, G. L. Hager, and M. Sanchez-Cespedes. 2012. The tumour suppressor and chromatin-remodelling factor BRG1 antagonizes myc activity and promotes cell differentiation in human cancer. *EMBO Molecular Medicine* 4 (7) (Jul): 603-16.

Ross, J. S., K. Wang, R. N. Al-Rohil, T. Nazeer, C. E. Sheehan, G. A. Otto, J. He, et al. 2014. Advanced urothelial carcinoma: Next-generation sequencing reveals diverse genomic alterations and targets of therapy. *Modern Pathology : An Official Journal of the United States and Canadian Academy of Pathology, Inc* 27 (2) (Feb): 271-80.

Rothman, N., M. Garcia-Closas, N. Chatterjee, N. Malats, X. Wu, J. D. Figueroa, F. X. Real, et al. 2010. A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. *Nature Genetics* 42 (11) (Nov): 978-84.

Rubio, E. D., D. J. Reiss, P. L. Welcsh, C. M. Disteche, G. N. Filippova, N. S. Baliga, R. Aebersold, J. A. Ranish, and A. Krumm. 2008. CTCF physically links cohesin to chromatin. *Proceedings of the National Academy of Sciences of the United States of America* 105 (24) (Jun 17): 8309-14.

Saal, L. H., P. Johansson, K. Holm, S. K. Gruvberger-Saal, Q. B. She, M. Maurer, S. Koujak, et al. 2007. Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity. *Proceedings of the National Academy of Sciences of the United States of America* 104 (18) (May 1): 7564-9.

Sabichi, A., A. Keyhani, N. Tanaka, J. Delacerda, I. L. Lee, C. Zou, J. H. Zhou, W. F. Benedict, and H. B. Grossman. 2006. Characterization of a panel of cell lines derived from urothelial neoplasms: Genetic alterations, growth in vivo and the relationship of adenoviral mediated gene transfer to coxsackie adenovirus receptor expression. *The Journal of Urology* 175 (3 Pt 1) (Mar): 1133-7.

Sajesh, B. V., B. J. Guppy, and K. J. McManus. 2013. Synthetic genetic targeting of genome instability in cancer. *Cancers* 5 (3) (Jun 24): 739-61.

Samanic, C., M. Kogevinas, M. Dosemeci, N. Malats, F. X. Real, M. Garcia-Closas, C. Serra, et al. 2006. Smoking and bladder cancer in spain: Effects of tobacco type, timing, environmental tobacco smoke, and gender. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 15 (7) (Jul): 1348-54.

Samlowski, W. E., S. A. Leachman, M. Wade, P. Cassidy, P. Porter-Gill, L. Busby, R. Wheeler, et al. 2005. Evaluation of a 7-day continuous intravenous infusion of decitabine: Inhibition of promoter-specific and global genomic DNA methylation. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* 23 (17) (Jun 10): 3897-905.

Sant, M., T. Aareleid, F. Berrino, M. Bielska Lasota, P. M. Carli, J. Faivre, P. Grosclaude, et al. 2003. EUROCARE-3: Survival of cancer patients diagnosed 1990-94--results and commentary. *Annals of Oncology : Official Journal of the European Society for Medical Oncology / ESMO* 14 Suppl 5 : v61-118.

Sarkis, A. S., G. Dalbagni, C. Cordon-Cardo, Z. F. Zhang, J. Sheinfeld, W. R. Fair, H. W. Herr, and V. E. Reuter. 1993. Nuclear overexpression of p53 protein in transitional cell bladder carcinoma: A marker for disease progression. *Journal of the National Cancer Institute* 85 (1) (Jan 6): 53-9.

Schaafsma, H. E., F. C. Ramaekers, G. N. van Muijen, E. B. Lane, I. M. Leigh, H. Robben, A. Huijsmans, E. C. Ooms, and D. J. Ruiter. 1990. Distribution of cytokeratin polypeptides in human transitional cell carcinomas, with special emphasis on changing expression patterns during tumor progression. *The American Journal of Pathology* 136 (2) (Feb): 329-43.

Schlessinger, J. 2000. Cell signaling by receptor tyrosine kinases. *Cell* 103 (2) (Oct 13): 211-25.

Schottenfeld, David., Fraumeni,Joseph F.,. 2006. *Cancer epidemiology and prevention*. Oxford; New York: Oxford University Press.

Seront, E., A. Pinto, C. Bouzin, L. Bertrand, J. P. Machiels, and O. Feron. 2013. PTEN deficiency is associated with reduced sensitivity to mTOR inhibitor in human bladder cancer through the unhampered feedback loop driving PI3K/Akt activation. *British Journal of Cancer* 109 (6) (Sep 17): 1586-92.

Seshagiri, S., E. W. Stawiski, S. Durinck, Z. Modrusan, E. E. Storm, C. B. Conboy, S. Chaudhuri, et al. 2012. Recurrent R-spondin fusions in colon cancer. *Nature* 488 (7413) (Aug 30): 660-4.

Shain, A. H., C. P. Giacomini, K. Matsukuma, C. A. Karikari, M. D. Bashyam, M. Hidalgo, A. Maitra, and J. R. Pollack. 2012. Convergent structural alterations define SWItch/Sucrose NonFermentable (SWI/SNF) chromatin remodeler as a central tumor suppressive complex in pancreatic cancer. *Proceedings of the National Academy of Sciences of the United States of America* 109 (5) (Jan 31): E252-9.

Shariat, S. F. 2007. Selecting patients for immediate cystectomy. *Reviews in Urology* 9 (4) (Fall): 239-41.

Siegel, R., J. Ma, Z. Zou, and A. Jemal. 2014. Cancer statistics, 2014. *CA: A Cancer Journal for Clinicians* 64 (1) (Jan-Feb): 9-29.

Signoretti, S., M. M. Pires, M. Lindauer, J. W. Horner, C. Grisanzio, S. Dhar, P. Majumder, et al. 2005. P63 regulates commitment to the prostate cell lineage. *Proceedings of the National Academy of Sciences of the United States of America* 102 (32) (Aug 9): 11355-60.

Silverman, D.T., S.S. Devesa, L.E. Moore, and N. Rothman. Bladder cancer. In: Schottenfeld D, Fraumeni JF Jr, editors. Cancer Epidemiology and Prevention. New York: Oxford University Press; 2006. pp. 1101–1127.

Singhal, N., J. Graumann, G. Wu, M. J. Arauzo-Bravo, D. W. Han, B. Greber, L. Gentile, M. Mann, and H. R. Scholer. 2010. Chromatin-remodeling components of the BAF complex facilitate reprogramming. *Cell* 141 (6) (Jun 11): 943-55.

Sjodahl, G., M. Lauss, K. Lovgren, G. Chebil, S. Gudjonsson, S. Veerla, O. Patschan, et al. 2012. A molecular taxonomy for urothelial carcinoma. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research* 18 (12) (Jun 15): 3377-86.

Snijders, A. M., R. Segraves, S. Blackwood, D. Pinkel, and D. G. Albertson. 2004. BAC microarray-based comparative genomic hybridization. *Methods in Molecular Biology (Clifton, N.J.)* 256 : 39-56.

Solomon, D. A., J. S. Kim, J. Bondaruk, S. F. Shariat, Z. F. Wang, A. G. Elkahloun, T. Ozawa, et al. 2013. Frequent truncating mutations of STAG2 in bladder cancer. *Nature Genetics* 45 (12) (Dec): 1428-30.

Solomon, D. A., T. Kim, L. A. Diaz-Martinez, J. Fair, A. G. Elkahloun, B. T. Harris, J. A. Toretsky, et al. 2011. Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science (New York, N.Y.)* 333 (6045) (Aug 19): 1039-43.

Spruck, C. H.,3rd, P. F. Ohneseit, M. Gonzalez-Zulueta, D. Esrig, N. Miyao, Y. C. Tsai, S. P. Lerner, C. Schmutte, A. S. Yang, and R. Cote. 1994. Two molecular pathways to transitional cell carcinoma of the bladder. *Cancer Research* 54 (3) (Feb 1): 784-8.

Stark, J. M., and M. Jasin. 2003. Extensive loss of heterozygosity is suppressed during homologous repair of chromosomal breaks. *Molecular and Cellular Biology* 23 (2) (Jan): 733-43.

Stenzl, A., N. C. Cowan, M. De Santis, M. A. Kuczyk, A. S. Merseburger, M. J. Ribal, A. Sherif, J. A. Witjes, and European Association of Urology (EAU). 2011. Treatment of muscle-invasive and

metastatic bladder cancer: Update of the EAU guidelines. *European Urology* 59 (6) (Jun): 1009-18.

Stratton, M. R., P. J. Campbell, and P. A. Futreal. 2009. The cancer genome. *Nature* 458 (7239) (Apr 9): 719-24.

Streppel, M. M., S. Lata, M. DelaBastide, E. A. Montgomery, J. S. Wang, M. I. Canto, A. M. Macgregor-Das, et al. 2014. Next-generation sequencing of endoscopic biopsies identifies ARID1A as a tumor-suppressor gene in barrett's esophagus. *Oncogene* 33 (3) (Jan 16): 347-57.

Svatek, R. S., B. K. Hollenbeck, S. Holmang, R. Lee, S. P. Kim, A. Stenzl, and Y. Lotan. 2014. The economics of bladder cancer: Costs and considerations of caring for this disease. *European Urology* (Jan 21).

Sylvester, R. J., A. P. van der Meijden, W. Oosterlinck, J. A. Witjes, C. Bouffioux, L. Denis, D. W. Newling, and K. Kurth. 2006. Predicting recurrence and progression in individual patients with stage ta T1 bladder cancer using EORTC risk tables: A combined analysis of 2596 patients from seven EORTC trials. *European Urology* 49 (3) (Mar): 466,5; discussion 475-7.

Tahara, T., E. Yamamoto, P. Madireddi, H. Suzuki, R. Maruyama, W. Chung, J. Garriga, et al. 2014. Colorectal carcinomas with CpG island methylator phenotype 1 frequently contain mutations in chromatin regulators. *Gastroenterology* 146 (2) (Feb): 530,38.e5.

Talaska, G., A. Z. al-Juburi, and F. F. Kadlubar. 1991. Smoking related carcinogen-DNA adducts in biopsy samples of human urinary bladder: Identification of N-(deoxyguanosin-8-yl)-4-aminobiphenyl as a major adduct. *Proceedings of the National Academy of Sciences of the United States of America* 88 (12) (Jun 15): 5350-4.

Tank, P. W., T. R. Gest, W. E. Burkel, W. Lippincott Williams & Wilkins. 2009. *Lippincott Williams & Wilkins atlas of anatomy*. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.

Taylor, C. F., F. M. Platt, C. D. Hurst, H. H. Thygesen, and M. A. Knowles. 2014. Frequent inactivating mutations of STAG2 in bladder cancer are associated with low tumour grade and stage and inversely related to chromosomal copy number changes. *Human Molecular Genetics* 23 (8) (Apr 15): 1964-74.

Thompson, E. R., M. A. Doyle, G. L. Ryland, S. M. Rowley, D. Y. Choong, R. W. Tothill, H. Thorne, et al. 2012. Exome sequencing identifies rare deleterious mutations in DNA repair genes FANCC and BLM as potential breast cancer susceptibility alleles. *PLoS Genetics* 8 (9) (Sep): e1002894.

Tilki, D., M. Brausi, R. Colombo, C. P. Evans, Y. Fradet, H. M. Fritsche, S. P. Lerner, A. Sagalowsky, S. F. Shariat, and B. H. Bochner. 2013. Lymphadenectomy for bladder cancer at the time of radical cystectomy. *European Urology* 64 (2) (Aug): 266-76.

Tomlinson, D. C., O. Baldo, P. Harnden, and M. A. Knowles. 2007. FGFR3 protein expression and its relationship to mutation status and prognostic variables in bladder cancer. *The Journal of Pathology* 213 (1) (Sep): 91-8.

Vallot, C., N. Stransky, I. Bernard-Pierrot, A. Herault, J. Zucman-Rossi, E. Chapeaublanc, D. Vordos, et al. 2011. A novel epigenetic phenotype associated with the most aggressive pathway of bladder tumor progression. *Journal of the National Cancer Institute* 103 (1) (Jan 5): 47-60.

van der Meijden, A. P., R. J. Sylvester, W. Oosterlinck, W. Hoeltl, A. V. Bono, and EORTC Genito-Urinary Tract Cancer Group. 2003. Maintenance bacillus calmette-guerin for ta T1 bladder tumors is not associated with increased toxicity: Results from a european organisation for research and treatment of cancer genito-urinary group phase III trial. *European Urology* 44 (4) (Oct): 429-34.

van Hemelrijck, M., A. Thorstenson, P. Smith, J. Adolfsson, and O. Akre. 2013. Risk of in-hospital complications after radical cystectomy for urinary bladder carcinoma: Population-based follow-up study of 7608 patients. *BJU International* 112 (8) (Dec): 1113-20.

van Rhijn, B. W., T. H. van der Kwast, A. N. Vis, W. J. Kirkels, E. R. Boeve, A. C. Jobsis, and E. C. Zwarthoff. 2004. FGFR3 and P53 characterize alternative genetic pathways in the pathogenesis of urothelial cell carcinoma. *Cancer Research* 64 (6) (Mar 15): 1911-4.

Vazquez, M., V. de la Torre, and A. Valencia. 2012. Chapter 14: Cancer genome analysis. *PLoS Computational Biology* 8 (12): e1002824.

Vineis, P., and R. Pirastu. 1997. Aromatic amines and cancer. *Cancer Causes & Control : CCC* 8 (3) (May): 346-55.

Volkmer, J. P., D. Sahoo, R. K. Chin, P. L. Ho, C. Tang, A. V. Kurtova, S. B. Willingham, et al. 2012. Three differentiation states risk-stratify bladder cancer into distinct subtypes. *Proceedings of the National Academy of Sciences of the United States of America* 109 (6) (Feb 7): 2078-83.

von der Maase, H., L. Sengelov, J. T. Roberts, S. Ricci, L. Dogliotti, T. Oliver, M. J. Moore, A. Zimmermann, and M. Arning. 2005. Long-term survival results of a randomized trial comparing gemcitabine plus cisplatin, with methotrexate, vinblastine, doxorubicin, plus cisplatin in patients with bladder cancer. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* 23 (21) (Jul 20): 4602-8.

Wang, D. D., Y. B. Chen, K. Pan, W. Wang, S. P. Chen, J. G. Chen, J. J. Zhao, et al. 2012. Decreased expression of the ARID1A gene is associated with poor prognosis in primary gastric cancer. *PloS One* 7 (7): e40364.

Wang, D. S., K. Rieger-Christ, J. M. Latini, A. Moinzadeh, J. Stoffel, J. A. Pezza, K. Saini, J. A. Libertino, and I. C. Summerhayes. 2000. Molecular analysis of PTEN and MXI1 in primary bladder carcinoma. *International Journal of Cancer.Journal International Du Cancer* 88 (4) (Nov 15): 620-5.

Wang, K., J. Kan, S. T. Yuen, S. T. Shi, K. M. Chu, S. Law, T. L. Chan, et al. 2011. Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nature Genetics* 43 (12) (Oct 30): 1219-23.

Welch, J. S., T. J. Ley, D. C. Link, C. A. Miller, D. E. Larson, D. C. Koboldt, L. D. Wartman, et al. 2012. The origin and evolution of mutations in acute myeloid leukemia. *Cell* 150 (2) (Jul 20): 264-78.

Wendt, K. S., K. Yoshida, T. Itoh, M. Bando, B. Koch, E. Schirghuber, S. Tsutsumi, et al. 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451 (7180) (Feb 14): 796-801.

Werner, H. M., A. Berg, E. Wik, E. Birkeland, C. Krakstad, K. Kusonmano, K. Petersen, et al. 2013. ARID1A loss is prevalent in endometrial hyperplasia with atypia and low-grade endometrioid carcinomas. *Modern Pathology : An Official Journal of the United States and Canadian Academy of Pathology, Inc* 26 (3) (Mar): 428-34.

Wiegand, K. C., S. P. Shah, O. M. Al-Agha, Y. Zhao, K. Tse, T. Zeng, J. Senz, et al. 2010. ARID1A mutations in endometriosis-associated ovarian carcinomas. *The New England Journal of Medicine* 363 (16) (Oct 14): 1532-43.

Williams, S. V., C. D. Hurst, and M. A. Knowles. 2013. Oncogenic FGFR3 gene fusions in bladder cancer. *Human Molecular Genetics* 22 (4) (Feb 15): 795-803.

Williamson, M. P., P. A. Elder, M. E. Shaw, J. Devlin, and M. A. Knowles. 1995. p16 (CDKN2) is a major deletion target at 9p21 in bladder cancer. *Human Molecular Genetics* 4 (9) (Sep): 1569-77.

Wilsker, D., A. Patsialou, S. D. Zumbrun, S. Kim, Y. Chen, P. B. Dallas, and E. Moran. 2004. The DNA-binding properties of the ARID-containing subunits of yeast and mammalian SWI/SNF complexes. *Nucleic Acids Research* 32 (4) (Feb 24): 1345-53.

Wilson, B. G., and C. W. Roberts. 2011. SWI/SNF nucleosome remodellers and cancer. *Nature Reviews.Cancer* 11 (7) (Jun 9): 481-92.

Witjes, J. A. 2004. Bladder carcinoma in situ in 2003: State of the art. *European Urology* 45 (2) (Feb): 142-6.

Wu, J. N., and C. W. Roberts. 2013. ARID1A mutations in cancer: Another epigenetic tumor suppressor? *Cancer Discovery* 3 (1) (Jan): 35-43.

Wu, R. C., T. L. Wang, and IeM Shih. 2014. The emerging roles of ARID1A in tumor suppression. *Cancer Biology & Therapy* 15 (6) (Jun 1): 655-64.

Wu, X., Y. Ye, L. A. Kiemeney, P. Sulem, T. Rafnar, G. Matullo, D. Seminara, et al. 2009. Genetic variation in the prostate stem cell antigen gene PSCA confers susceptibility to urinary bladder cancer. *Nature Genetics* 41 (9) (Sep): 991-5.

Wynder, E. L., A. Augustine, G. C. Kabat, and J. R. Hebert. 1988. Effect of the type of cigarette smoked on bladder cancer risk. *Cancer* 61 (3) (Feb 1): 622-7.

Xiao, T., J. Wallace, and G. Felsenfeld. 2011. Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Molecular and Cellular Biology* 31 (11) (Jun): 2174-83.

Yamamoto, S., H. Tsuda, M. Takano, S. Tamai, and O. Matsubara. 2012. Loss of ARID1A protein expression occurs as an early event in ovarian clear-cell carcinoma development and frequently coexists with PIK3CA mutations. *Modern Pathology : An Official Journal of the United States and Canadian Academy of Pathology, Inc* 25 (4) (Apr): 615-24.

Yan, J., S. Roy, A. Apolloni, A. Lane, and J. F. Hancock. 1998. Ras isoforms vary in their ability to activate raf-1 and phosphoinositide 3-kinase. *The Journal of Biological Chemistry* 273 (37) (Sep 11): 24052-6.

Yang, A., M. Kaghad, Y. Wang, E. Gillett, M. D. Fleming, V. Dotsch, N. C. Andrews, D. Caput, and F. McKeon. 1998. P63, a P53 homolog at 3q27-29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities. *Molecular Cell* 2 (3) (Sep): 305-16.

Yang, X., F. Lay, H. Han, and P. A. Jones. 2010. Targeting DNA methylation for epigenetic therapy. *Trends in Pharmacological Sciences* 31 (11) (Nov): 536-46.

Yeager, T. R., S. DeVries, D. F. Jarrard, C. Kao, S. Y. Nakada, T. D. Moon, R. Bruskewitz, et al. 1998. Overcoming cellular senescence in human cancer pathogenesis. *Genes & Development* 12 (2) (Jan 15): 163-74.

Zang, Z. J., I. Cutcutache, S. L. Poon, S. L. Zhang, J. R. McPherson, J. Tao, V. Rajasegaran, et al. 2012. Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nature Genetics* 44 (5) (May): 570-4.

Zeegers, M. P., F. E. Tan, E. Dorant, and P. A. van Den Brandt. 2000. The impact of characteristics of cigarette smoking on urinary tract cancer risk: A meta-analysis of epidemiologic studies. *Cancer* 89 (3) (Aug 1): 630-9.

Zhang, X., Q. Sun, M. Shan, M. Niu, T. Liu, B. Xia, X. Liang, et al. 2013. Promoter hypermethylation of ARID1A gene is responsible for its low mRNA expression in many invasive breast cancers. *PloS One* 8 (1): e53931.

Zhang, Z. T., J. Pak, E. Shapiro, T. T. Sun, and X. R. Wu. 1999. Urothelium-specific expression of an oncogene in transgenic mice induced the formation of carcinoma in situ and invasive transitional cell carcinoma. *Cancer Research* 59 (14) (Jul 15): 3512-7.

Zhao, L., and P. K. Vogt. 2008. Helical domain and kinase domain mutations in p110alpha of phosphatidylinositol 3-kinase induce gain of function by different mechanisms. *Proceedings of the National Academy of Sciences of the United States of America* 105 (7) (Feb 19): 2652-7.

Zieger, K., L. Dyrskjot, C. Wiuf, J. L. Jensen, C. L. Andersen, K. M. Jensen, and T. F. Orntoft. 2005. Role of activating fibroblast growth factor receptor 3 mutations in the development of bladder tumors. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research* 11 (21) (Nov 1): 7709-19.

Annex I. Supplementary tables

Supplementary Table 1. Characteristics of the patients included in the discovery and prevalence screens. E: Sample included in the exome sequencing screen. P: Sample included in HaloPlex prevalence screen. A: Sample included in the ARID1A mutational screen. NA: Information not available. S: Smoker. NS: Non-smoker. ES: Ex-smoker.

Sample	Age	Gender	Stage/ Grade	Smoking status	Study of inclusion
62	60	Male	T1G3	S	E
114	76	Female	T2G3	NS	E
116	83	Male	T1G2	S	E
179	56	Male	TaG1	S	E
193	71	Male	T2G3	S	E
251	84	Female	TaG2	NS	E
310	46	Male	T1G3	NS	E
331	47	Male	TaG1	S	E
343	81	Female	T1G3	NS	E
413	58	Male	TaG3	S	E
1418	77	Female	TaG1	NS	E
Esp66	68	Male	T1G3	S	EA
64	95	Female	T1G2	NS	EP
188	53	Male	T1G3	S	EP
274	35	Male	T1G2	NS	EP
313	73	Male	T1G3	S	EP
451	78	Male	TaG2	S	EP
11	66	Male	T1G3	NA	P
47	78	Male	T1G3	NA	P
76	64	Male	T1G3	S	P
80	73	Male	TaG2	NA	P
88	83	Male	TaG1	ES	P
102	75	Male	TaG1	ES	P
121	80	Male	TaG3	ES	P
131	77	Male	T1G3	NA	P
147	51	Male	TaG1	S	P
181	59	Male	TaG1	NA	P
207	59	Male	T1G2	S	P
248	82	Male	T1G3	NA	P
272	36	Male	T1G3	ES	P
275	82	Male	T2G3	NA	P
280	74	Male	T2G3	S	P
281	86	Male	TaG2	ES	P
304	76	Male	TaG2	NS	P
308	85	Male	TaG3	NA	P
311	57	Male	TaG3	S	P
312	84	Male	TaG3	ES	P

325	73	Male	TisG2	NA	P
339	53	Male	TaG2	NA	P
344	60	Male	TaG1	NA	P
358	75	Male	T2G3	NA	P
375	79	Male	T1G2	NS	P
395	82	Female	TaG3	ES	P
402	73	Male	T1G3	NA	P
411	82	Male	TisG3	ES	P
414	71	Male	T1G3	ES	P
416	74	Male	TaG1	S	P
424	72	Male	TaG3	ES	P
427	51	Male	TaG1	ES	P
432	70	Male	T1G3	S	P
433	66	Male	T1G3	S	P
435	83	Male	T1G3	ES	P
438	81	Male	T1G3	ES	P
439	74	Male	T1G3	ES	P
450	81	Male	T1G3	ES	P
453	95	Male	TaG2	NS	P
473	69	Female	TaG1	NS	P
479	58	Male	TaG1	NA	P
489	63	Female	NA	NS	P
497	77	Male	T1G3	S	P
508	66	Male	TaG2	NA	P
512	74	Female	T1G3	NA	P
533	53	Male	TaG2	S	P
562	83	Male	T1G2	ES	P
597	83	Male	TaG1	ES	P
630	89	Female	T2G3	NS	P
639	48	Male	T1G2	S	P
718	69	Male	T1G3	ES	P
776	66	Male	T1G3	NA	P
811	80	Male	T2G3	ES	P
818	76	Female	TaG2	NS	P
823	76	Male	T1G3	ES	P
Esp104	45	Male	TaG2	ES	P
Esp21	74	Male	TaG1	S	PA
Esp27	77	Male	T2G3	ES	PA
Esp44	69	Male	TaG3	S	P
Esp86	56	Male	T2G3	S	PA
Esp55	67	Male	T3G3	NA	A
Esp64	64	Male	TaG1	NA	A
Esp46	72	Male	TaG3	NA	A

Esp95	77	Male	TaG1	NA	A
Esp49	68	Male	TaG2	NA	A
Esp41	79	Male	T3G3	NA	A
Esp69	75	Male	T3G3	NA	A
Esp37	71	Male	TaG3	NA	A
Esp52	73	Male	TaG2	NA	A
Esp29	71	Male	TaG3	NA	A
Esp61	58	Male	T1G2	NA	A
5	54	Male	TaG1	NA	A
19	72	Male	TaG1	NA	A
28	74	Male	TaG2	NA	A
31	54	Male	TaG1	NA	A
34	50	Male	TaG1	NA	A
35	56	Male	TaG1	NA	A
43	NA	Male	T4G3	NA	A
12	71	Male	T1G3	NA	A
789	51	Male	TaG1	NA	A
791	82	Male	T1G2	NA	A
793	80	Male	TaG3	NA	A
796	76	Male	T1G3	NA	A
798	60	Male	T1G3	NA	A
800	64	Male	T1G3	NA	A
803	53	Male	T1G3	NA	A
811b	83	Male	TaG1	NA	A
814	76	Female	T2G3	NA	A
818b	59	Male	T1G2	NA	A
826	95	Female	T1G2	NA	A
828	86	Male	TaG3	NA	A
830	73	Male	TaG2	NA	A
833	51	Male	T1G2	NA	A
835	84	Female	TaG2	NA	A
842	81	Male	T1G3	NA	A
851	81	Male	T2G3	NA	A
856	56	Male	TaG1	NA	A
870	46	Male	T1G3	NA	A
871	57	Male	TaG3	NA	A
874	84	Male	TaG3	NA	A
875	73	Male	T1G3	NA	A
877	76	Male	TaG2	NA	A
557	47	Male	TaG1	NA	A
559	82	Female	T1G3	NA	A
563	72	Male	TaG3	NA	A
742	53	Male	TaG2	NA	A
816	NA	NA	NA	NA	A
744	80	Male	T1G2	NA	A

Supplementary Table 2. **Characteristics of the patients with UBC for which ARID1A expression was assessed.** NA: Information not available.

		N (%)
Total		84
Age	Mean (SD)	66.4 (9.7)
Gender	Male	74 (88.1)
	Female	10 (11.9)
Stage/Grade	TaG1	21 (25.0)
	TaG2	18 (21.4)
	TaG3	7 (8.3)
	T1G2	NA
	T1G3	12 (14.3)
	> T2	26 (31.0)
Tumor size	< 3 cm	27 (32.1)
	≥ 3 cm	17 (20.2)
	Unknown	40 (47.7)
Multiplicity	Single	47 (56.0)
	Multiple	29 (34.5)
	Unknown	8 (9.5)
Treatment	TUR alone	12 (14.4)
	TUR+endovesical chemo	15 (17.6)
	TUR+BCG	30 (35.8)
	Cystectomy	10 (11.9)
	TUR+BCG+endov chemo	NA
	TUR+cystectomy	NA
	Systemic Chemotherapy	8 (9.6)
	Radiotherapy	5 (5.9)
	Others	4 (4.8)
	Missing	NA

Supplementary Table 3. Characteristics of the patients with UBC for which STAG2 protein expression was assessed. NA: Not applicable.

	NMIUBC		MIUBC	
	N	**%**	**N**	**%**
Total	480	100	182	100
Gender				
Male	427	89	160	87.9
Female	53	11	22	12.1
Age				
≤60	118	24.6	39	21.4
61-70	190	39.6	74	40.7
>70	172	35.8	69	37.9
Area				
Barcelona	84	17.5	33	18.1
Valles-Bages	84	17.5	33	18.1
Alicante	43	9	11	6
Tenerife	101	21	31	17
Asturias	168	35	74	40.7
Tumor number				
Solitary	303	67.6	122	73.9
Multiple	145	32.4	43	26.1
Tumor site				
1 site	298	63.5	84	46.9
>1 site	171	36.5	95	53.1
Tumor size				
≤3cm	254	52.9	40	22
>3cm	96	20	57	31.3
Unknown	130	27.1	85	46.7
Histological grade				
Benign (PUNLMP)	1	0.2	NA	NA
G1	156	32.5	NA	NA
G2	173	36	14	7.7
G3	150	31.2	168	92.3
T stage				
Ta	359	74.8	NA	NA
T1	121	25.2	NA	NA
T2	NA	NA	92	50.5
T3	NA	NA	48	26.4
T4	NA	NA	42	23.1
TG category				
TaG1	157	32.7	NA	NA
TaG2	167	34.8	NA	NA
TaG3	35	7.3	NA	NA
T1G2	6	1.2	NA	NA

T1G3	115	24	NA	NA
T2G2	NA	NA	6	3.3
T2G3	NA	NA	86	47.3
T3G2	NA	NA	5	2.7
T3G3	NA	NA	43	23.6
T4G2	NA	NA	3	1.6
T4G3	NA	NA	39	21.4
Risk groups				
Low-risk (TaG1, TaG2)	324	67.5	NA	NA
High-risk (T1G2,TaG3, T1G3)	156	32.5	NA	NA
Treatment				
Others	27	5.6	51	28.2
RTU	188	39.2	NA	NA
RTU + BCG	149	31	NA	NA
RTU + BCG + Intravesical Chemoth.	16	3.3	NA	NA
RTU + Intrav. Chemotherapy	100	20.8	NA	NA
Cystectomy	NA	NA	58	32
Cystectomy+chemotherapy	NA	NA	31	17.1
Chemotherapy	NA	NA	20	11
Radiotherapy $\pm$ Chemotherapy	NA	NA	13	7.2
Intravesical treatment	NA	NA	8	4.4
Number of recurrences				
0	327	68.1	NA	NA
1 - 2	110	22.9	NA	NA
> 2	43	9	NA	NA
Nodal status				
0	NA	NA	119	65.4
1	NA	NA	45	24.7
2	NA	NA	18	9.9
Metastasis				
M0	NA	NA	146	80.2
M1	NA	NA	24	13.2
Mx	NA	NA	12	6.6

Supplementary Table 4. **Genes included in the prevalence screen and reason for inclusion**.

Gene Name	Pathway/Reason for Inclusion
APEX1	Base Excision Repair
APEX2	Base Excision Repair
BPTF	Chromatin remodelling
BRCA2	Fanconi anemia
BRIP1	Fanconi anemia
BTBD12	Fanconi anemia
ERCC2	DNA repair
ESCO1	Cohesin/Chromosome segregation
ESCO2	Cohesin/Chromosome segregation
ESPL1	Cohesin/Chromosome segregation
EXO1	DNA repair
FANCA	Fanconi anemia
FANCB	Fanconi anemia
FANCC	Fanconi anemia
FANCE	Fanconi anemia
FANCF	Fanconi anemia
FANCG	Fanconi anemia
FANCI	Fanconi anemia
FANCL	Fanconi anemia
HRAS	Involved in UBC
kiaa0892	Cohesin/Chromosome segregation
KRAS	Involved in UBC
LIG3	Base Excision Repair
MBD4	Base Excision Repair
MLH1	DNA repair
MPG	Base Excision Repair
MSH4	DNA repair
MTMR15/FAN1	DNA repair
MUTYH	Base Excision Repair
NEIL3	Base Excision Repair
NOTCH1	DNA repair
NOTCH2	DNA repair
OGG1	Base Excision Repair
PARP1	DNA repair
PDS5A	Cohesin/Chromosome segregation
PDS5B	Cohesin/Chromosome segregation
PMS1	DNA repair
PNKP	Base Excision Repair
POLD1	DNA repair
POLE	DNA repair
POLN	DNA repair
PTEN	Involved in UBC
RAD18	DNA repair

RAD21	Cohesin/Chromosome segregation
RAD50	DNA repair
RAD51C	Fanconi anemia
REC8	Cohesin/Chromosome segregation
SMC1A	Cohesin/Chromosome segregation
SMC3	Cohesin/Chromosome segregation
SMUG1	Base Excision Repair
STAG1	Cohesin/Chromosome segregation
STAG3	Cohesin/Chromosome segregation
TOP2A	Cohesin/Chromosome segregation
TP53BP1	DNA repair
UNG	Base Excision Repair
WAPAL	Cohesin/Chromosome segregation
XRCC1	Base Excision Repair
ARHGAP5	Recurrent in Discovery screen
ARID1A	Recurrent in Discovery screen
ASXL1	Recurrent in Discovery screen
ASXL2	Recurrent in Discovery screen
ATM	DNA repair
BCLAF1	Recurrent in Discovery screen
BCOR	Recurrent in Discovery screen
BRAF	Recurrent in Discovery screen
CASP5	Recurrent in Discovery screen
CHD6	Recurrent in Discovery screen
CHEK2	DNA repair
CREBBP	Recurrent in Discovery screen
CSE1L	Recurrent in Discovery screen
DHX32	Recurrent in Discovery screen
DISP1	Recurrent in Discovery screen
EIF2C4	Recurrent in Discovery screen
ELF3	Recurrent in Discovery screen
EP300	Recurrent in Discovery screen
ERBB3	Involved in UBC
FANCD2	Fanconi anemia
FANCM	Fanconi anemia
FBXW7	Cohesin/Chromosome segregation
FGFR3	Involved in UBC
ILK	Recurrent in Discovery screen
IREB2	Recurrent in Discovery screen
KDM6A/UTX	Recurrent in Discovery screen
KHSRP	Recurrent in Discovery screen
KIF2B	Cohesin/Chromosome segregation
MAP3K7	Recurrent in Discovery screen
MAP4K4	Recurrent in Discovery screen
MAPK14	Recurrent in Discovery screen

MAPK8IP3	Recurrent in Discovery screen
MLL	Recurrent in Discovery screen
MLL2	Recurrent in Discovery screen
MLL3	Recurrent in Discovery screen
MYCBP2	Recurrent in Discovery screen
MYSM1	Recurrent in Discovery screen
NBN	DNA repair
NCOR1	Recurrent in Discovery screen
NIPBL	Cohesin/Chromosome segregation
NLRP5	Recurrent in Discovery screen
NRAS	Involved in UBC
NUP93	Recurrent in Discovery screen
PALB2	Fanconi anemia
PIK3CA	Involved in UBC
RAD51	DNA repair
RAD54B	DNA repair
RB1	Involved in UBC
SMC1B	Cohesin/Chromosome segregation
SMCHD1	Recurrent in Discovery screen
STAG2	Cohesin/Chromosome segregation
TP53	Involved in UBC
WHSC1L1	Recurrent in Discovery screen

Supplementary Table 5. **Coverage metrics for samples included in the exome sequencing screen**. N: Normal leukocyte DNA. T: Tumor DNA. tROI: Targeted Regions of Interest. C10: Percentage of bases covered by more than 10 reads. C15: Percentage of bases covered by more than 15 reads.

Sample	tROI	Specificity	Enrichment	C10	C15	Mean_cov	Median_cov
418 N	46830265	73.27	148.10	84.30	79.68	83.79	58
418 T	46830265	73.09	146.72	84.26	79.38	72.66	53
331 N	46830265	72.10	139.65	82.09	76.36	62.02	44
331 T	46830265	72.99	146.01	81.47	75.41	53.95	40
179 N	50570842	71.19	123.42	93.14	88.48	73.30	55
179 T	50570842	72.40	131.02	95.49	92.26	93.51	70
251 N	46830265	73.03	146.31	84.16	79.46	78.64	56
251 T	46830265	72.83	144.81	84.02	79.22	75.60	54
451 N	50570842	71.05	122.63	96.43	94.05	115.31	88
451 T	50570842	71.66	126.34	94.83	91.51	100.73	73
413 N	46830265	72.33	141.19	81.78	75.89	58.39	42
413 T	46830265	72.07	139.42	81.64	75.58	55.65	40
064 N	50570842	71.72	126.68	94.98	91.63	93.70	71
064 T	50570842	72.09	129.01	95.00	91.62	98.14	71
274 N	50570842	70.16	117.44	95.53	92.38	92.04	71
274 T	50570842	70.77	120.93	95.18	92.09	97.83	75
116 N	46830265	71.92	138.34	83.22	78.21	75.83	52
116 T	46830265	71.08	132.78	80.82	74.53	55.59	39
313 N	50570842	70.49	119.35	95.55	92.52	101.34	75
313 T	50570842	69.17	112.08	92.96	88.34	76.79	56
188 N	50570842	70.20	117.66	95.05	91.60	91.58	68
188 T	50570842	70.10	117.10	94.96	91.65	98.67	71
062 N	50570842	69.43	113.46	96.61	94.13	112.33	84
062 T	50570842	68.97	111.05	94.53	90.75	85.49	64
Esp66 N	46830265	74.39	156.92	79.87	73.77	50.97	39
Esp66 T	46830265	43.47	41.54	66.93	62.20	74.59	33
310 N	46830265	69.19	121.34	83.69	78.80	75.83	53
310 T	46830265	72.56	142.89	83.35	78.12	68.83	49
343 N	50570842	70.85	121.41	94.13	90.19	83.90	62
343 T	50570842	68.46	108.43	93.21	88.77	79.80	57
114 N	46830265	73.09	146.76	83.69	79.02	83.91	57
114 T	46830265	71.80	137.55	83.04	78.47	88.49	59
193 N	46830265	71.20	133.56	86.09	81.24	66.85	50
193 T	46830265	69.63	123.88	84.89	79.33	59.07	44
Mean		70.55	127.82	88.14	83.72	80.44	58.03
SD		5.00	19.88	7.10	8.00	16.87	13.77

Median		71.43	127.85	85.49	80.46	79.22	56.50
Min		43.47	41.54	66.93	62.20	50.97	33.00
Max		74.39	156.92	96.61	94.13	115.31	88.00

Mean N		71.51	131.42	88.84	84.55	82.34	60.29
SD N		1.43	13.28	6.33	7.40	17.77	14.10
Median N		71.20	126.68	86.09	81.24	83.79	57.00
Min N		69.19	113.46	79.87	73.77	50.97	39.00
Max N		74.39	156.92	96.61	94.13	115.31	88.00

Mean T		69.60	124.21	87.45	82.90	78.55	55.76
SD T		6.89	24.72	7.93	8.71	16.24	13.47
Median T		71.66	129.01	84.89	79.38	76.79	56.00
Min T		43.47	41.54	66.93	62.20	53.95	33.00
Max T		73.09	146.72	95.49	92.26	100.73	75.00
P-value		0.48	0.45	0.43	0.41	0.61	0.47

Supplementary Table 6. **Sanger sequencing validation of alterations identified by exome sequencing.**

Sample	Gene Symbol	Genomic Position	Base change	Exon junction mutation	Sanger validation
	BCLAF1	6:136599393	G>C		Yes/Germline
	BCOR	X:39930312	C>A		Yes
114	*CAP2*	6:17514088	G>A		Yes
	NBEAL1	2:203977847	C>A		Yes
	NCOA6	20:33328965	G>C		Yes
	ACOT12	5:80640707	G>A	Yes	Yes
	APAF1	12:99109247	C>A		Yes
	APAF1	12:99119192	G>C	Yes	Yes
	ARHGEF38	4:106587443	T>A		Yes
	ARSD	X:2836238	G>A		Yes
	ATP11B	3:182631645	G>A	Yes	Yes
	BCAS3	17:59067579	C>T		Yes
	BCLAF1	6:136599393	G>C		Yes/Germline
	C20orf43	20:55093380	C>T		Yes
	CASP5	11:104877890	G>C		Yes
	DCAF4L1	4:41984354	T>A		Yes
	DGKG	3:186024771	A>G	Yes	Yes
	DHX32	10:127530345	C>T		Yes
	DPYD	1:98164907	C>A	Yes	Yes
	ENOX1	13:43986126	G>T		Yes
	ERBB3	12:56482552	G>C		Yes
	ERBB3	12:56488226	G>A		Yes
116	*FANCD2*	3:10084271	C>T		Yes
	FBXO15	18:71749182	T>G		Yes
	GGTLC2	22:22989270	G>C		Yes
	HECW1	7:43351361	G>A	Yes	Yes
	IREB2	15:78770698	C>T		Yes
	MAP4K4	2:102476144	G>A		Yes
	MAPK8IP3	16:1817211	C>G		Yes
	NBN	8:90955475	A>G	Yes	Yes
	OBSCN	1:228487711	G>C		Yes
	OR8D4	11:123777880	G>A		Not validated
	RELL2	5:141017844	C>T		Yes
	SLITRK2	X:144905201	A>G		Yes
	TMEM126A	11:85366652	G>C		Yes
	WDR67	8:124153010	G>C		Yes
	WWC1	5:167850687	C>G		Yes
	ZNF382	19:37118108	A>G		Yes
	ZNF513	2:27600816	G>A		Yes
	ZNF645	X:22291997	C>G		Yes

	EP300	22:41545917	T>+T		Yes
	NUP98	11:3735079	T>C		Yes
179	*PIK3CA*	3:178916944	A>G		Yes
	TBC1D19	4:26640413	A>T		Yes
	UBE3B	12:109935628	T>C		Yes
188	*RBM10*	X:47034489	C>T	Yes	Yes
	ADCYAP1R1	7:31146175	C>G		Yes
	AKNAD1	1:109391616	G>A		Yes
	ANK2	4:114161701	C>T		Yes
	ART4	12:14993738	G>A		Yes
	ASXL1	20:31024236	G>A		Yes
	ATP2A3	17:3853842	C>T		Yes
	AUTS2	7:70246606	G>A		Yes
	CEP70	3:138224288	G>A		Yes
	COL27A1	9:116930982	A>C		Yes
	CREBBP	16:3790439	G>C		Yes
	CREM	10:35477169	G>A		Yes
	CSE1L	20:47689123	C>T		Yes
	DNAH2	17:7674239	G>A		Yes
	DOCK3	3:51352413	G>A	Yes	Yes
	FAM83D	20:37580315	C>T		Yes
	FBXO43	8:101153886	C>T		Yes
	FBXW7	4:153247184	G>A		Yes
	GABRA1	5:161277884	G>A		Yes
193	*GABRP*	5:170222429	G>A	Yes	Yes
	GHR	5:42689058	G>A		Yes
	IMPA1	8:82593792	C>T		Yes
	KCTD10	12:109893973	C>T		Yes
	KDM6A	X:44894233	A>-----	Yes	Yes
	NCOA4	10:51582942	C>T	Yes	Yes
	NSD1	5:176683980	G>C		Yes
	NSD1	5:176684158	G>C	Yes	Yes
	NUP93	16:56782202	C>T		Yes
	PDZD2	5:32010451	G>T		Yes
	PHIP	6:79679800	T>C		Yes
	PIK3CA	3:178921553	T>A		Yes
	PRAMEF11	1:12884817	C>A		Yes
	PTPRD	9:8389266	A>G		Yes
	PUM1	1:31426622	A>T		Yes
	RGMA	15:93588794	C>T		Yes
	STAG2	X:123204998	G>A	Yes	Yes
	TIMM23B	10:51582942	C>T	Yes	Yes
	TMEM117	12:44782322	C>T		Yes
	WHSC1L1	8:38186970	T>C		Not validated

251	*ARHGAP5*	14:32562174	A>G		Yes
	CHD6	20:40161768	C>T		Yes
	CREBBP	16:3808977	C>T	Yes	Not validated
	CTTN	11:70253616	C>G	Yes	Yes
	CUL3	2:225422486	G>C		Yes
	EHBP1	2:63176088	G>A		Yes
	ELF3	1:201984363	A>--------		Yes
	KCNK2	1:215256745	C>T		Yes
	KIAA1370	15:52902548	T>A		Yes
	LY75	2:160755489	C>T	Yes	Yes
	LY75-CD302	2:160755489	C>T	Yes	Yes
	MAP2	2:210560516	C>A		Yes
	MTIF2	2:55481309	G>A		Yes
	MYSM1	1:59127132	C>T		Yes
	NUP133	1:229600575	C>G		Yes
	PIK3CA	3:178916876	G>A		Yes
	RHOA	3:49405915	G>T		Yes
	SAMD9L	7:92762797	C>T		Yes
	TMEM173	5:138858001	C>T		Yes
	WHSC1L1	8:38162239	T>C		Yes
274	*GNAS*	20:57480483	C>T		Yes
310	*KDM5B*	1:202715021	T>C		Yes
	RSC1A1	1:15986544	G>A		Yes
	AC022098.3	19:14269152	C>G		Yes
	ADD3	10:111876075	G>+CA		Yes
	BRMS1L	14:36334139	A>G	Yes	Yes
	CACNA1D	3:53757641	C>T		Yes
	DHX32	10:127555638	C>T		Yes
	KDM6A	X:44969442	T>-		Yes
	LPHN1	19:14269152	C>G		Yes
	MLL2	12:49418394	C>A		Yes
	NLRP5	19:56539603	G>A		Yes
	RAD54B	8:95448825	G>C		Yes
	STAG2	X:123200023	A>T	Yes	Yes
	ZNF428	19:44111720	C>T		Yes
413	*ABHD13*	13:108881846	A>T		Yes
	ADAM32	8:39114795	G>T		Yes
	ATM	11:108143533	G>A		Yes
	BRAF	7:140481441	T>C		Yes
	CHD9	16:53262979	G>C		Yes
	DISP1	1:223168260	C>G		Yes
	FGFR3	4:1806092	A>T		Yes
	GOT1	10:101166543	G>A		Yes
	HBP1	7:106820415	C>T		Yes

	ICAM5	19:10402180	T>C		Yes
	IQGAP2	5:75969287	G>A		Yes
	LIMK1	7:73535377	G>A	Yes	Yes
	MAPK14	6:36041477	T>-		Yes
	MLL2	12:49420607	G>A		Yes
	MLL3	7:151851423	A>T		Yes
	MYH7	14:23894113	C>G		Yes
	NIPBL	5:37008194	A>G	Yes	Yes
	NME7	1:169292395	C>A		Yes
	NPR1	1:153653686	G>C		Yes
	PRDX4	X:23704428	G>A		Yes
	PRKCH	14:61952298	G>A		Yes
	PROCR	20:33764632	C>G		Yes/Germline
	RAD51	15:41001249	G>T		Yes
	SMC1B	22:45795226	C>T		Yes
	STXBP5	6:147588215	C>G	Yes	Yes
	TRPM2	21:45819236	G>A		Yes
	UBR4	1:19474549	G>A		Yes
	XKR4	8:56270260	G>T		Yes
	ZNF462	9:109688261	G>C		Yes
	ZNF716	7:57528942	C>T		Yes
418	*ASXL2*	2:25978922	G>T		Yes
	ASXL2	2:25982401	C>A		Yes
	ASXL2	2:25982439	C>T		Yes
	CHD4	12:6707456	C>G		Yes
	CHEK2	22:29095903	C>G		Yes
	CTNND1	11:57571085	T>C	Yes	Yes
	FGFR3	4:1806099	A>G		Yes
	GAB2	11:77936174	C>T		Yes
	GPR98	5:90079045	G>A		Yes
	ILK	11:6631209	A>C		Yes
	KHSRP	19:6418562	G>A		Yes
	MTHFR	1:11863014	C>G		Yes
	MYEF2	15:48450972	C>A		Yes
	MYSM1	1:59125701	C>T		Yes
	NBEA	13:36220438	C>T		Yes
	PIK3CA	3:178936091	G>A		Yes
	SMCHD1	18:2703807	C>-----		Yes
	STAG2	X:123196830	C>T		Yes
	ULK4	3:41757012	G>C		Yes
451	*CR2*	1:207647031	C>G		Yes
	DNAH3	16:21147804	G>A		Yes
	FLT3	13:28626737	C>G		Yes
	LIPG	18:47101866	G>T		Yes

	MADCAM1	19:504757	C>T		Yes
	MAGI3	1:114193672	G>A		Yes
	MDN1	6:90372684	C>G	Yes	Yes
	PABPC1L	20:43585027	G>A	Yes	Yes
	RAPGEF2	4:160279265	G>A	Yes	Yes
	SLFN13	17:33767862	C>T		Yes
	SULT1E1	4:70710007	G>A		Yes
	TAF5	10:105146985	C>G		Not validated
	TOMM34	20:43585027	G>A	Yes	Yes
	TP53	17:7578394	T>A		Yes
064	*ADRA1A*	8:26722392	A>G		Yes
	HCN3	1:155254496	C>G		Yes
	HCN3	1:155254514	C>T		Yes
	KDM4D	11:94731144	A>G		Yes
	NF1	17:29701298	C>T		Yes
	PHKA2	X:18938281	G>A		Yes
	PPP4R4	14:94732129	A>C	Yes	Yes
	SLC38A9	5:54931377	C>G		Yes
	TP53	17:7578211	C>A		Yes
Esp66	*BCHE*	3:165547615	C>A		Yes
	CDKAL1	6:20846406	C>T		Yes
	DSP	6:7568676	C>T		Yes
	EIF2C1	1:36367916	C>A		Yes
	FANK1	10:127668859	C>T		Yes/Germline
	KRT35	17:39633349	G>C		Yes/Germline

Supplementary Table 7. **Depth of coverage metrics for samples included in the HaloPlex targeted resequencing screen**. B: Normal DNA from blood (note that some of the normal samples were analyzed in pools). Targeted Regions of Interest: 485,052 for all cases. C10: Percentage of bases covered by more than 10 reads. C15: Percentage of bases covered by more than 15 reads. C30: Percentage of bases covered by more than 30 reads.

Sample ID	C10	C15	C30	Mean_Cov	Median_Cov
011	96,3324	95,4584	93,6745	1243,0404	796
047	78,8775	74,1898	63,7771	180,2932	62
076	67,9502	62,3383	51,6780	128,5843	33
080	97,1898	96,5674	94,7686	1003,3752	584
088	95,8081	94,9366	92,6789	869,1008	576
102	96,7278	96,0992	94,7134	2397,8991	1612
121	74,4968	69,0241	57,4992	139,7130	45
131	67,0973	61,5705	50,2105	90,3180	31
147	59,3483	53,4959	42,6482	121,8625	19
181	95,0601	94,0584	91,7625	1466,4578	748
207	68,8048	63,1681	50,4797	99,1697	31
248	91,4506	89,4271	84,3835	348,6409	205
272	81,3408	79,1422	74,5403	1055,5705	409
275	91,5250	89,6211	85,0459	420,4908	240
280	94,6218	93,4240	90,2578	563,3873	384
281	98,0085	97,6215	96,6791	1114,3599	871
304	93,7998	92,6672	89,9083	714,7739	426
308	98,4198	98,1895	97,5576	1655,5875	1340
311	90,6233	88,5495	83,5791	367,8674	203
312	89,4003	86,7150	80,1974	270,9763	141
325	95,3553	94,1901	90,9488	1872,3075	407
339	92,3464	90,5330	86,0780	404,6770	242
344	95,3923	94,2909	91,6786	592,1638	419
358	88,5757	86,8579	83,2255	1088,7451	624
375	89,2195	87,5088	83,7972	818,3309	363
395	68,6599	62,7133	50,6954	139,0895	32
402	94,2790	92,8678	89,3725	519,6964	341
411	75,6098	71,1460	61,9812	236,9898	63
414	92,1145	90,5528	86,6157	428,4106	271
416	93,2820	92,1153	89,3964	836,3758	569
424	98,4558	98,1835	97,4966	1721,2199	1168
427	90,5862	89,1739	85,8279	563,4743	319
432	93,5714	92,4573	89,4450	904,5724	471
433	95,7586	95,0055	92,6975	991,9533	626
435	98,3018	98,0355	97,3285	1863,1357	1297
438	92,4130	91,1350	88,4054	1163,1750	579
439	86,4235	84,1081	78,7858	412,4881	169
450	98,3449	98,0579	97,3073	1510,1652	1172
453	87,7809	85,9040	82,2732	1152,5467	619
473	97,5596	96,9624	95,4737	1278,7138	866
479	97,8388	97,4003	96,1990	1364,3534	927
489	85,9063	84,1211	79,9995	862,9771	511
497	96,8403	96,1814	94,0182	861,7447	473
508	76,5159	73,9426	68,8038	498,6970	175

512	98,4193	98,1647	97,3902	1447,0394	1104
533	98,1171	97,8545	97,0343	1960,5307	1451
562	85,3407	83,2694	79,0606	909,6617	464
597	98,4781	98,2517	97,6135	1896,5084	1360
630	96,3420	95,4727	92,8521	544,5656	334
639	86,8787	85,0843	80,9697	1020,0184	564
718	95,9755	95,0230	91,9501	540,5156	339
776	84,3699	82,0792	77,6836	1108,3872	425
811	81,8597	79,7403	76,1815	1128,2851	477
818	80,7761	78,7495	75,0732	1195,5729	535
823	84,1394	82,0275	78,0755	1068,0800	540
Esp104	96,8694	96,2470	94,5410	808,2399	550
Esp21	97,0915	96,4119	94,5414	689,7494	450
Esp27	96,3373	95,5992	93,4296	793,9407	464
Esp44	95,9969	95,0335	91,9681	391,3194	254
Esp86	97,9153	97,5009	96,5457	1078,6478	756
011B_088B_432B	98,6670	98,4480	97,8518	1709,0528	1187
076B_181B_439B	98,6834	98,4833	97,9771	2643,8851	1896
121B_047B_207B	96,2810	95,5685	94,1243	5299,8896	3401
131B_Esp21B_188B_438B	98,7220	98,5383	98,0582	7927,1577	4818
133B_304B_402B	96,4136	95,5986	93,9875	2015,3740	1136
248B_275B_311B_312B	98,7449	98,5160	98,0788	3273,4870	2208
280B_450B_424B_597B	98,9393	98,8193	98,4756	5283,2169	3861
281B_512B_433B	98,9028	98,7403	98,2757	3247,6938	2231
313B_339B_375B_451B	98,4593	98,1971	97,5205	2334,3425	1502
344B_414B_080B	95,7403	95,0572	93,3211	2157,1065	1381
395B_411B_147B	98,5288	98,2575	97,5580	2974,4779	2050
435B_416B_427B	98,9535	98,7803	98,3674	3206,8752	2431
473B_102B_497B	98,6789	98,4781	97,9800	2434,4296	1741
479B_325B_308B	98,8333	98,6597	98,2416	3305,0814	2307
630B_640B_644B_533B	98,8562	98,7146	98,3517	3384,1756	2386
718B_064B_274B	98,7337	98,4958	97,8714	3084,5725	2298
Esp27B_Esp104B_Esp44B_Esp86B	98,9045	98,7496	98,3437	3028,3054	2129
272-B	86,8670	85,0296	81,2766	1519,3046	820
358-B	86,5769	84,7008	81,2907	1454,8019	816
453-B	85,4754	83,1513	79,2080	977,7152	517
489-B	84,4332	82,3246	77,6954	1122,0791	581
508-B	79,1746	77,0606	72,7320	1016,9928	324
562-B	86,0217	83,8419	79,8595	923,8382	503
639-B	83,0971	80,7876	76,5724	1132,2674	461
818-B	75,6667	73,4816	68,6609	1113,7520	236
776-B	84,3233	82,2386	78,1648	1373,3648	620
811-B	82,3314	80,2877	75,9494	1182,5767	449
823-B	80,9810	78,8153	74,6514	1103,7262	409

Supplementary Table 8. **Recurrently altered pathways in the discovery and prevalence screens**.

Summary. Y: yes; N: no.

Pathway	Non synonymous (NS)	NS damaging	NS recurrent	NS recurrent damaging
Chromatin modification	Y	Y	Y	Y
Apoptosis	Y	Y	Y	Y
DNA repair and DNA damage response	Y	Y	Y	Y
Cell cycle	Y	Y	N	N
Sister chromatid cohesion	Y	Y	N	N

Gene Ontology Biological Processes enrichment (Including non-synonymous mutations and mutations in exon junctions, n=908)

GO ID	p-value (Adjusted by FDR)
positive regulation of transcription, DNA-dependent	0.00153
negative regulation of gene-specific transcription from RNA polymerase II promoter	0.00153
blood coagulation	0.00153
response to DNA damage stimulus	0.00153
chromatin modification	0.00198
negative regulation of smoothened signaling pathway	0.00200
regulation of transcription, DNA-dependent	0.00200
heart development	0.00271
N-terminal peptidyl-lysine acetylation	0.00399
androgen receptor signaling pathway	0.00480
regulation of heart contraction	0.00509
cellular response to ionizing radiation	0.00647
positive regulation of protein import into nucleus, translocation	0.00647
axon guidance	0.00647
smooth muscle cell differentiation	0.00819
base conversion or substitution editing	0.00819
maintenance of organ identity	0.00819
cell-cell adhesion	0.00833
base-excision repair	0.00855
Rho protein signal transduction	0.00855
cell-matrix adhesion	0.00952
neuron projection development	0.01061
protein transport	0.01184
ATP catabolic process	0.01282
positive regulation of gene-specific transcription from RNA polymerase II promoter	0.01863
skin development	0.01863
activation of adenylate cyclase activity	0.01946
lung morphogenesis	0.02004
interspecies interaction between organisms	0.02150
transcription, DNA-dependent	0.02150
cell cycle	0.02230
negative regulation of transcription from RNA polymerase II promoter	0.02248

cellular protein localization	0.02248
ion transport	0.02248
regulation of angiogenesis	0.02438
gene silencing by RNA	0.02438
positive regulation of canonical Wnt receptor signaling pathway	0.02438
axonogenesis	0.02503
cell morphogenesis	0.02641
establishment or maintenance of cell polarity	0.02641
mesenchymal to epithelial transition involved in metanephros morphogenesis	0.02641
cell redox homeostasis	0.02641
response to calcium ion	0.02689
negative regulation of transcription, DNA-dependent	0.03008
integrin-mediated signaling pathway	0.03008
cell cycle arrest	0.03008
histone H3-K4 methylation	0.03008
estrogen receptor signaling pathway	0.03008
negative regulation of translation	0.03072
stress fiber assembly	0.03072
replicative senescence	0.03072
regulation of transcription factor activity	0.03072
double-strand break repair via homologous recombination	0.03072
cell morphogenesis involved in differentiation	0.03072
positive regulation of osteoblast differentiation	0.03072
regulation of epithelial cell differentiation	0.03072
cellular response to UV	0.03172
nerve growth factor receptor signaling pathway	0.03748
leukocyte migration	0.03779
platelet activation	0.03779
negative regulation of gene expression	0.03993
detection of calcium ion	0.04034
positive regulation of transcription from RNA polymerase II promoter	0.04160
branching involved in ureteric bud morphogenesis	0.04311
negative regulation of DNA binding	0.04311
positive regulation of axon extension	0.04311
branched chain family amino acid catabolic process	0.04311
MAPKKK cascade	0.04390
in utero embryonic development	0.04390
regulation of insulin secretion	0.04494
response to hormone stimulus	0.04591
small GTPase mediated signal transduction	0.04591
regulation of vasoconstriction	0.04591
vesicle-mediated transport	0.04591
regulation of JNK cascade	0.04591
2-oxoglutarate metabolic process	0.04591
DNA damage response, signal transduction resulting in induction of apoptosis	0.05098
digestive tract development	0.05098
response to organic nitrogen	0.05098
transcription from RNA polymerase I promoter	0.05098
regulation of glucose transport	0.05098
transcription initiation from RNA polymerase I promoter	0.05098
positive regulation of reactive oxygen species metabolic process	0.05098
sister chromatid cohesion	0.05189
cytoskeletal anchoring at plasma membrane	0.05189

midgut development	0.05189
regulation of mitotic metaphase/anaphase transition	0.05189
positive regulation of protein tyrosine kinase activity	0.05189
oocyte development	0.05189
activation of phospholipase C activity	0.05436
cell cycle checkpoint	0.05436
determination of left/right symmetry	0.05599
protein phosphorylation	0.05971
Wnt receptor signaling pathway	0.05971
positive regulation of dendrite morphogenesis	0.05971
chromatin silencing	0.05971
negative regulation of apoptosis	0.05971
Ras protein signal transduction	0.05971
sensory perception of light stimulus	0.05971
photoreceptor cell maintenance	0.05971
positive regulation of gene expression	0.06513
mitosis	0.06782
DNA damage checkpoint	0.06782
protein localization	0.06782
negative regulation of macrophage derived foam cell differentiation	0.06843
cell division	0.06843
adult heart development	0.06843
mRNA stabilization	0.06843
regulation of small GTPase mediated signal transduction	0.06843
lens morphogenesis in camera-type eye	0.06843
cellular response to insulin stimulus	0.06861
response to wounding	0.06889
mitotic cell cycle	0.07027
neuron apoptosis	0.07065
response to morphine	0.07065
negative regulation of cell growth	0.07339
cell migration	0.07491
response to inorganic substance	0.07491
skeletal muscle tissue development	0.07491
insulin-like growth factor receptor signaling pathway	0.07491
protein heterooligomerization	0.07491
hair follicle morphogenesis	0.07491
negative regulation of osteoclast differentiation	0.07491
potassium ion transport	0.07491
protein localization to nucleus	0.07491
nervous system development	0.07594
response to hydrogen peroxide	0.07897
positive regulation of MAPKKK cascade	0.07897
G2/M transition of mitotic cell cycle	0.08044
3'-phosphoadenosine 5'-phosphosulfate metabolic process	0.08044
protein ubiquitination	0.08044
positive regulation of erythrocyte differentiation	0.08044
negative regulation of angiogenesis	0.08044
activation of JUN kinase activity	0.08044
blood vessel development	0.08044
face morphogenesis	0.08044
cilium assembly	0.08044
regulation of translational initiation	0.08044

phosphatidylinositol phosphorylation	0.08044
spindle assembly	0.08044
regulation of type I interferon-mediated signaling pathway	0.08455
cellular process	0.08455
histone acetylation	0.08455
mRNA transport	0.08455
positive regulation of peptidyl-serine phosphorylation	0.08845
potassium ion transmembrane transport	0.08845
induction of apoptosis by intracellular signals	0.08845
cAMP-mediated signaling	0.08845
synapse organization	0.08845
phospholipid catabolic process	0.08845
regulation of interferon-gamma-mediated signaling pathway	0.08845
vasculogenesis	0.08845
forebrain development	0.09038
DNA repair	0.09058
vitamin metabolic process	0.09065
regulation of cell migration	0.09316
hexose transport	0.09316
cell differentiation	0.09488
G1 phase of mitotic cell cycle	0.09539
positive regulation of epithelial to mesenchymal transition	0.09539
cellular response to organic cyclic compound	0.09539
response to salt stress	0.09539
positive regulation of cell-substrate adhesion	0.09539
positive regulation of Wnt receptor signaling pathway	0.09539
positive regulation of branching involved in ureteric bud morphogenesis	0.09539
protein secretion	0.09539

Gene Ontology Biological Processes enrichment (Including non-synonymous mutations and mutations in exon junctions that were recurrent, n=68)

GO ID	p-value (Adjusted by FDR)
positive regulation of apoptosis	0.00417
positive regulation of neuron apoptosis	0.00417
cell adhesion	0.00444
chromatin modification	0.00707
liver development	0.00865
Ras protein signal transduction	0.00865
heart development	0.00898
negative regulation of gene-specific transcription from RNA polymerase II promoter	0.00898
response to hypoxia	0.01054
positive regulation of gene-specific transcription from RNA polymerase II promoter	0.01054
regulation of transcription, DNA-dependent	0.01054
cell migration	0.01054
interspecies interaction between organisms	0.01108
positive regulation of transcription, DNA-dependent	0.01201
negative regulation of apoptosis	0.01719
homophilic cell adhesion	0.01719
response to DNA damage stimulus	0.01719
positive regulation of transcription from RNA polymerase II promoter	0.02126

negative regulation of transcription, DNA-dependent	0.02302
blood coagulation	0.02569
negative regulation of cell proliferation	0.04140
transcription, DNA-dependent	0.04865

Gene Ontology Biological Processes enrichment (Including non-synonymous mutations and mutations in exon junctions that were predicted to be damaging, n=565)

GO ID	p-value (Adjusted by FDR)
chromatin modification	0.00009
androgen receptor signaling pathway	0.00159
response to DNA damage stimulus	0.00159
regulation of transcription, DNA-dependent	0.00159
N-terminal peptidyl-lysine acetylation	0.00159
cellular response to ionizing radiation	0.00159
maintenance of organ identity	0.00325
cell-cell adhesion	0.00411
cell cycle arrest	0.00435
transcription, DNA-dependent	0.00435
negative regulation of gene-specific transcription from RNA polymerase II promoter	0.00508
base-excision repair	0.00508
heart development	0.00508
blood coagulation	0.00630
cell redox homeostasis	0.00876
positive regulation of transcription, DNA-dependent	0.00944
histone H3-K4 methylation	0.00949
cell morphogenesis	0.01062
cellular response to UV	0.01109
negative regulation of transcription, DNA-dependent	0.01164
Rho protein signal transduction	0.01178
regulation of epithelial cell differentiation	0.01178
stress fiber assembly	0.01178
replicative senescence	0.01178
cell morphogenesis involved in differentiation	0.01178
cell-matrix adhesion	0.01214
positive regulation of gene-specific transcription from RNA polymerase II promoter	0.01214
ion transport	0.01341
double-strand break repair via homologous recombination	0.01364
2-oxoglutarate metabolic process	0.01364
protein phosphorylation	0.01364
ATP catabolic process	0.01416
leukocyte migration	0.01504
skin development	0.01504
DNA damage response, signal transduction resulting in induction of apoptosis	0.01504
interspecies interaction between organisms	0.01504
signal transduction	0.01504
positive regulation of reactive oxygen species metabolic process	0.01504
axon guidance	0.01504
regulation of JNK cascade	0.01518
negative regulation of gene expression	0.01587
negative regulation of transcription from RNA polymerase II promoter	0.01592
positive regulation of MAPKKK cascade	0.01629

regulation of angiogenesis	0.01629
photoreceptor cell maintenance	0.01629
regulation of small GTPase mediated signal transduction	0.01629
MAPKKK cascade	0.01671
regulation of mitotic metaphase/anaphase transition	0.01671
regulation of heart contraction	0.01671
sister chromatid cohesion	0.01671
DNA damage checkpoint	0.01743
protein transport	0.01793
Ras protein signal transduction	0.01883
neuron apoptosis	0.01913
sensory perception of light stimulus	0.01913
positive regulation of dendrite morphogenesis	0.01913
negative regulation of translation	0.02140
integrin-mediated signaling pathway	0.02224
mRNA stabilization	0.02224
lens morphogenesis in camera-type eye	0.02224
induction of apoptosis by intracellular signals	0.02224
in utero embryonic development	0.02224
histone acetylation	0.02539
cellular process	0.02539
positive regulation of transcription from RNA polymerase II promoter	0.02656
protein localization	0.02656
cellular response to insulin stimulus	0.02830
Wnt receptor signaling pathway	0.02890
protein ubiquitination	0.02973
estrogen receptor signaling pathway	0.02973
G1 phase of mitotic cell cycle	0.02973
smooth muscle cell differentiation	0.02973
apoptosis	0.03047
cell aging	0.03227
negative regulation of smoothened signaling pathway	0.03367
transcription initiation from RNA polymerase I promoter	0.03430
cell cycle checkpoint	0.03430
transcription from RNA polymerase I promoter	0.03430
neuron projection development	0.03430
response to hormone stimulus	0.03430
positive regulation of gene expression	0.03823
mitotic cell cycle	0.03978
forebrain development	0.03978
patterning of blood vessels	0.03978
skeletal muscle tissue development	0.03978
branched chain family amino acid catabolic process	0.03991
positive regulation of axon extension	0.03991
negative regulation of DNA binding	0.03991
small GTPase mediated signal transduction	0.03991
nerve growth factor receptor signaling pathway	0.04227
metanephros development	0.04534
protein autophosphorylation	0.04689
digestive tract development	0.04862
histone methylation	0.04862
gastrulation with mouth forming second	0.04862
positive regulation of osteoblast differentiation	0.04862

transcription initiation from RNA polymerase II promoter	0.04862
retinoic acid receptor signaling pathway	0.04862
cell differentiation	0.04929
response to calcium ion	0.05282
negative regulation of neuron apoptosis	0.05458
negative regulation of apoptosis	0.05579
positive regulation of neuron apoptosis	0.05807
nervous system development	0.06068
double-strand break repair	0.06146
branching involved in ureteric bud morphogenesis	0.06146
establishment or maintenance of cell polarity	0.06476
regulation of mitotic cell cycle	0.06476
positive regulation of peptidyl-serine phosphorylation	0.06492
actin cytoskeleton organization	0.06515
regulation of cell migration	0.06907
response to morphine	0.06941
regulation of long-term neuronal synaptic plasticity	0.06941
endocytosis	0.07078
response to organic nitrogen	0.07143
positive regulation of canonical Wnt receptor signaling pathway	0.07143
glucose metabolic process	0.07204
ventricular cardiac muscle tissue morphogenesis	0.07289
regulation of multicellular organism growth	0.07289
hair follicle morphogenesis	0.07289
protein heterooligomerization	0.07298
regulation of protein localization	0.07701
regulation of translational initiation	0.07701
positive regulation of multicellular organism growth	0.07701
transcription elongation from RNA polymerase I promoter	0.07701
spindle assembly	0.07701
extracellular matrix organization	0.08074
G2/M transition of mitotic cell cycle	0.08119
positive regulation of muscle cell differentiation	0.08305
S phase of mitotic cell cycle	0.08322
intracellular signal transduction	0.08405
vesicle-mediated transport	0.08697
termination of RNA polymerase I transcription	0.08698
transcription from RNA polymerase II promoter	0.08698
negative regulation of neuron differentiation	0.08698
embryonic hindlimb morphogenesis	0.08698
activation of adenylate cyclase activity	0.09208
fertilization	0.09208
cell migration	0.09208
calcium ion transport	0.09208
response to organic cyclic compound	0.09804
response to glucocorticoid stimulus	0.09831
neuroprotection	0.09831
negative regulation of protein kinase activity	0.09942

Gene Ontology Biological Processes enrichment (Including n-synonymous mutations and mutations in exon junctions that were recurrent and predicted to be damaging, n=58)

GO ID	P-value (Adjusted by FDR)
positive regulation of neuron apoptosis	0.00195
positive regulation of apoptosis	0.00195
heart development	0.00383
Ras protein signal transduction	0.00383
cell adhesion	0.00383
regulation of transcription, DNA-dependent	0.00383
negative regulation of gene-specific transcription from RNA polymerase II promoter	0.00383
liver development	0.00383
response to hypoxia	0.00497
interspecies interaction between organisms	0.00497
positive regulation of gene-specific transcription from RNA polymerase II promoter	0.00497
chromatin modification	0.00767
homophilic cell adhesion	0.00937
positive regulation of transcription from RNA polymerase II promoter	0.00937
response to DNA damage stimulus	0.00937
negative regulation of apoptosis	0.00937
blood coagulation	0.01133
negative regulation of transcription, DNA-dependent	0.01133
negative regulation of cell proliferation	0.01913
positive regulation of transcription, DNA-dependent	0.01913
transcription, DNA-dependent	0.01913
cell differentiation	0.06563

KEGG pathways enrichment (Including n-synonymous mutations and mutations in exon junctions, n=908)

KEGG Pathway ID	p-value (Adjusted by FDR)
Regulation of actin cytoskeleton	0.02121
Adherens junction	0.02998
Cell cycle	0.09840

KEGG pathways enrichment (Including n-synonymous mutations and mutations in exon junctions that were recurrent, n=68)

KEGG Pathway ID	p-value (Adjusted by FDR)
Cell cycle	0.00005
Adherens junction	0.00718
Wnt signaling pathway	0.00718
Apoptosis	0.00946
Leukocyte transendothelial migration	0.01619
Jak-STAT signaling pathway	0.02880
Focal adhesion	0.04778
Regulation of actin cytoskeleton	0.04957
MAPK signaling pathway	0.07623

KEGG pathways enrichment (Including n-synonymous mutations and mutations in exon junctions that were predicted to be damaging, n=565)

KEGG Pathway ID	p-value (Adjusted by FDR)
Regulation of actin cytoskeleton	0.01436
Adherens junction	0.01436
Bacterial invasion of epithelial cells	0.03650
Tight junction	0.04867
Leukocyte transendothelial migration	0.04867
p53 signaling pathway	0.06443
Cell cycle	0.06578
Focal adhesion	0.06736
Toxoplasmosis	0.06736
MAPK signaling pathway	0.06736
Homologous recombination	0.06736
Pantothenate and CoA biosynthesis	0.07308

KEGG pathways enrichment (Including n-synonymous mutations and mutations in exon junctions that were recurrent and predicted to be damaging, n=56)

KEGG Pathway ID	p-value (Adjusted by FDR)
Cell cycle	0.00003
Adherens junction	0.00541
Wnt signaling pathway	0.00541
Apoptosis	0.00715
Leukocyte transendothelial migration	0.01234
Jak-STAT signaling pathway	0.02215
Focal adhesion	0.03717
Regulation of actin cytoskeleton	0.03869
MAPK signaling pathway	0.06028

Supplementary Table 9. Descriptive analysis of *ARID1A* mutations and expression in relationship with other UBC markers.

	N	Mean(SD)	Median(IQR)	P-value t-test/Anova	P-value MW/KW
ARID1A_Score ALL	84	156.8 (104.5)	180 (135)		
ARID1A_Score by Sup_Inv				**$9.7x10^{-6}$**	**$3.5x10^{-5}$**
Sup	58	189.8 (93.2)	200 (132.5)		
Invas	26	83.1 (91.1)	50 (180)		
ARID1A_Score by Risk Groups				**$9.9x10^{-6}$**	**$3.7x10^{-5}$**
Low	39	206.4 (80.2)	200 (120)		
High	19	155.8 (110.1)	180 (160)		
Invas	26	83.1 (91.1)	50 (180)		
ARID1A_Score by FGFR3_mutation				0.41	0.78
WT	17	181.8 (109.2)	200 (200)		
Muta	30	207 (81.3)	200 (120)		
FGFR3_Score ALL	77	57.3 (68.2)	30 (98)		
FGFR3_Score by Sup_Inv				0.46	0.28
Sup	54	61.3 (65.2)	47.5 (99.5)		
Invas	23	47.8 (75.5)	5 (73)		
FGFR3_Score by Risk Groups				**0.038**	**0.026**
Low	39	73.5 (67.2)	70 (117)		
High	15	29.7 (48.4)	0 (47.5)		
Invas	23	47.8 (75.5)	5 (73)		
FGFR3_Score by FGFR3_mutation				**$3.1x10^{-5}$**	**0.00032**
WT	15	19.3 (32.9)	0 (30)		
Muta	29	90 (67.4)	80 (110)		
p53_1801_Score ALL	34	65.5 (100.7)	14.5 (56.2)		
p53_1801_Score by Sup_Inv				**0.017**	0.21
Sup	22	28.3 (56.6)	12 (16.8)		
Invas	12	133.8 (128.2)	120 (258)		
p53_1801_Score by Risk Groups				0.05	0.32
Low	18	16.2 (17.5)	11 (15.8)		
High	4	82.5 (126.1)	30 (82.5)		
Invas	12	133.8 (128.2)	120 (258)		
p53_1801_Score by FGFR3_mutation				0.11	0.17
WT	3	5.3 (7.6)	2 (7)		
Muta	15	35.9 (67.6)	12 (34.5)		
p53_DO7_Score ALL	83	81.7 (102.6)	30 (108)		
p53_DO7_Score by Sup_Inv				**0.0042**	**0.029**
Sup	57	56.6 (82)	20 (48)		
Invas	26	136.9 (121.7)	120 (274.5)		
p53_DO7_Score by Risk Groups				**$6.3x10^{-5}$**	**0.0022**
Low	38	28.7 (45.6)	18 (20)		
High	19	112.3 (108.3)	60 (180)		
Invas	26	136.9 (121.7)	120 (274.5)		
p53_DO7_Score by FGFR3_mutation				**0.011**	**0.014**
WT	17	123.5 (115.7)	105 (216)		
Muta	30	38.8 (65.6)	20 (26)		

Pearson Correlations

	ARID1A	FGFR3	p53_1801	p53_DO7
ARID1A	1	0.208	-0.209	-0.189
FGFR3	0.208	1	-0.244	-0.249
p53_1801	-0.209	-0.244	1	0.979
p53_DO7	-0.189	-0.249	0.979	1

Pearson Correlations P-Value

	ARID1A	FGFR3	p53_1801	p53_DO7
ARID1A		0.070	0.236	0.088
FGFR3	0.070		0.186	**0.030**
p53_1801	0.236	0.186		**0.000**
p53_DO7	0.088	**0.030**	**0.000**	

Spearman Correlations

	ARID1A	FGFR3	p53_1801	p53_DO7
ARID1A	1	0.247	-0.207	-0.115
FGFR3	0.247	1	-0.157	-0.146
p53_1801	-0.207	-0.157	1	0.929
p53_DO7	-0.115	-0.146	0.929	1

Spearman Correlations P-Value

	ARID1A	FGFR3	p53_1801	p53_DO7
ARID1A		**0.030**	0.239	0.300
FGFR3	**0.0300**		0.400	0.208
p53_1801	0.239	0.400		**0.000**
p53_DO7	0.300	0.208	**0.000**	

Supplementary Table 10. **Survival analysis of UBC according to ARID1A expression.**

	Events N(%)	Censored N(%)	HR	95% CI	P-Value Cox PH	P-Value Log-Rank
Recurrence - All						
Low ARID1A	5(14.7)	29(85.3)	1 (Ref)			
High ARID1A	22(44)	28(56)	3,28	1.24 - 8.67	**0.017**	**0.011**
Recurrence - NMI						
Low ARID1A	5(31.2)	11(68.8)	1 (Ref)			
High ARID1A	22(52.4)	20(47.6)	1,93	0.73 - 5.11	0.184	0.177
Recurrence - Low Risk (NMI)						
Low ARID1A	2(25)	6(75)	1 (Ref)			
High ARID1A	16(51.6)	15(48.4)	2,62	0.6 - 11.42	0.199	0.182
Recurrence - High Risk (NMI)						
Low ARID1A	3(37.5)	5(62.5)	1 (Ref)			
High ARID1A	6(54.5)	5(45.5)	1,53	0.38 - 6.17	0.552	0.549
Progression - All						
Low ARID1A	10(29.4)	24(70.6)	1 (Ref)			
High ARID1A	8(16)	42(84)	0,48	0.19 - 1.21	0.12	0.112
Progression - NMI						
Low ARID1A	2(12.5)	14(87.5)	1 (Ref)			
High ARID1A	3(7.1)	39(92.9)	0,6	0.1 - 3.57	0.571	0.567
Progression - Low Risk (NMI)						
Low ARID1A	1(12.5)	7(87.5)	1 (Ref)			
High ARID1A	1(3.2)	30(96.8)	0,29	0.02 - 4.64	0.381	0.351
Progression - High Risk (NMI)						
Low ARID1A	1(12.5)	7(87.5)	1 (Ref)			
High ARID1A	2(18.2)	9(81.8)	1,37	0.12 - 15.17	0.795	0.794
Progression - MI						
Low ARID1A	8(44.4)	10(55.6)	1 (Ref)			
High ARID1A	5(62.5)	3(37.5)	1,53	0.5 - 4.69	0.456	0.453
Mortality - All						
Low ARID1A	7(20.6)	27(79.4)	1 (Ref)			
High ARID1A	7(14)	43(86)	0,58	0.2 - 1.65	0.307	0.301
Mortality - MI						
Low ARID1A	5(27.8)	13(72.2)	1 (Ref)			
High ARID1A	5(62.5)	3(37.5)	2,24	0.65 - 7.75	0.203	0.191

Supplementary Table 11. **Association between STAG2 expression and patient and tumor characteristics.**

		N	STAG2 ≤50 N(%)	STAG2 >50 N(%)	P-value (*)
Gender					0.828
	Male	595	176 (89.3)	419 (88.4)	
	Female	76	21 (10.7)	55 (11.6)	
Age					0.056
	≤60	157	58 (29.7)	99 (21.2)	
	61-70	264	74 (37.9)	190 (40.7)	
	>70	241	63 (32.3)	178 (38.1)	
Area					0.769
	Barcelona	117	36 (18.5)	81 (17.3)	
	Valles-Bages	117	38 (19.5)	79 (16.9)	
	Alicante	54	17 (8.7)	37 (7.9)	
	Tenerife	132	40 (20.5)	92 (19.7)	
	Asturias	242	64 (32.8)	178 (38.1)	
Tumor number					0.606
	Solitary	431	133 (71.1)	298 (68.7)	
	Multiple	190	54 (28.9)	136 (31.3)	
Tumor site					**0.011**
	1 site	387	130 (66.7)	257 (55.6)	
	>1 site	270	65 (33.3)	205 (44.4)	
Tumor size					**0.002**
	≤3cm	295	107 (54.3)	188 (39.7)	
	>3cm	155	40 (20.3)	115 (24.3)	
	Unknown	221	50 (25.4)	171 (36.1)	
Invasiveness					**$2.71x10^{-8}$**
	NMIBC	480	171 (87.7)	309 (66.2)	
	MIBC	182	24 (12.3)	158 (33.8)	
Risk group					**$4.29x10^{-11}$**
	Low-risk	324	134 (68.7)	190 (40.7)	
	High-risk	156	37 (19)	119 (25.5)	
	Invasive	182	24 (12.3)	158 (33.8)	
TG category					**$2.20x10^{-16}$**
	TaG1	156	83 (42.8)	73 (15.6)	
	TaG2	167	50 (25.8)	117 (25.1)	
	TaG3	35	21 (10.8)	14 (3)	
	T1G2	6	3 (1.5)	3 (0.6)	
	T1G3	115	13 (6.7)	102 (21.8)	
	T2G2	6	2 (1)	4 (0.9)	
	T2G3	86	8 (4.1)	78 (16.7)	
	T3G2	5	3 (1.5)	2 (0.4)	
	T3G3	43	2 (1)	41 (8.8)	

	T4G2	3	0 (0)	3 (0.6)	
	T4G3	39	9 (4.6)	30 (6.4)	
T stage					**$5.72x10^{-15}$**
	Ta	359	155 (79.5)	204 (43.7)	
	T1	121	16 (8.2)	105 (22.5)	
	T2	92	10 (5.1)	82 (17.6)	
	T3	48	5 (2.6)	43 (9.2)	
	T4	42	9 (4.6)	33 (7.1)	
Histological grade					**$1.96x10^{-15}$**
	G1	156	83 (42.8)	73 (15.6)	
	G2	187	58 (29.9)	129 (27.6)	
	G3	318	53 (27.3)	265 (56.7)	

(*) P-value of the chi-squared test

Supplementary Table 12. Association between ARID1A mutations and tumor/patient characteristics in all UBC studies.

Compared categories	Total number (mutated)	P (Fisher´s exact test or Chi2 test)
NMI	187 (27) = 14.4%	**0.02**
MI	258 (61) = 23.6%	
NMI-LR	120 (17) = 14.2%	**0.05**
NMI-HR	67 (10) = 14.9%	
MI	258 (61) = 23.6%	
Non-aggressive	74 (12) = 16.2%	0.4
Aggressive	371 (76) = 20.5%	
Females	84 (20) = 23.8%	0.2
Males	362 (61) = 16.9%	
<60	89 (12) = 13.5%	0.6
>60	226 (36) = 15.9%	
Never smoked	119 (23) = 19.3%	0.9
Smoke(d)	210 (43) = 20.5%	

Supplementary Table 13. Mutual exclusivity and co-occurrence analyses between *ARID1A* alterations and other genes in all UBC studies.

KRAS

		Mutant	Wild Type			
ARID1A	Mutant	5	19	24	p left	1
	Wild Type	6	119	125	p right	**0.02**
		11	138	149	OR	5.2

FGFR3

		Mutant	Wild Type			
ARID1A	Mutant	10	98	108	p left	**0.007**
	Wild Type	77	318	395	p right	1
		87	416	503	OR	**0.4**

STAG2

		Mutant	Wild Type			
ARID1A	Mutant	6	85	91	p left	**0.03**
	Wild Type	50	289	339	p right	1
		56	374	430	OR	**0.41**

TP53

		Mutant	Wild Type			
ARID1A	Mutant	47	61	108	p left	0.0
	Wild Type	148	247	395	p right	0.2
		195	308	503	OR	1.3

Supplementary Table 14. Association between STAG2 mutations and tumor/patient characteristics in all UBC studies.

Compared categories	Total number (mutated)	P (Fisher´s exact test or Chi2 test)
NMI	414 (109) = 26.3%	**<0.0001**
MI	348 (38) = 10.9%	
NMI-LR	228 (68) = 29.8%	**<0.0001**
NMI-HR	166 (36) = 21.7%	
MI	348 (38) = 10.9%	
Non-aggressive	189 (56) = 29.6%	**<0.0001**
Aggressive	591 (93) = 15.7%	
Females	198 (44) = 22.2%	**0.01**
Males	545 (78) = 14.3%	
<60	87 (18) = 20.7%	0.2
>60	220 (32) = 14.6%	
Never smoked	156 (18) = 11.5%	0.07
Smoke(d)	242 (45) = 18.6%	

Supplementary Table 15. Mutual exclusivity and co-occurrence analyses between STAG2 alterations and other genes in all UBC studies.

PIK3CA

		Mutant	Wild Type			
STAG2	Mutant	8	48	56	p left	0.4
	Wild Type	63	311	374	p right	0.7
		71	359	430	OR	0.8

KRAS

		Mutant	Wild Type			
STAG2	Mutant	1	13	14	p left	**0.0**
	Wild Type	8	92	100	p right	1
		9	105	114	OR	**0.005**

FGFR3

		Mutant	Wild Type			
STAG2	Mutant	23	33	56	p left	1
	Wild Type	49	325	374	p right	**0.0**
		72	358	430	OR	**4.6**

TP53

		Mutant	Wild Type			
STAG2	Mutant	16	29	45	p left	0.2
	Wild Type	124	162	286	p right	0.9
		140	191	331	OR	0.7

Annex II. Publications